Investigating factors that influence English L2+ perceptual vowel acquisition: modality, speaker, and feedback

Marc Jones

Acknowledgements

I would like to thank my friends and family for support and understanding I could not always be there.

I would also like to pass on my thanks and support for everybody that has supported me while preparing, and during my doctoral studies, including co-authors, editors, and colleagues and especially my supervisor.

Last, but by no means least, I would like to thank my students at Toyo University, without whom there would be nothing to actually write about.

Abstract

This thesis is based upon a selected 5 papers and articles themed around perceptual phonology acquisition and listening for English as a second/additional language (L2+). The articles are focused upon the use of Global Englishes and prestige varieties as a source of vowel acquisition (Jones & Blume, 2022) and as potential factors in listening difficulty (Jones 2024d); visual salience as a factor in vowel acquisition (Jones, 2024c), and the effect of feedback timing on vowel acquisition (Jones, 2024e). Additionally, a conference paper discusses potential methods for small sample analysis, particularly Bayesian methods (Jones 2024a), which uses Jones and Blume (2022) as an example, but which is also a rationale for the analysis methods in all three of the quantitative articles (Jones & Blume, 2022; Jones 2024c and 2024e).

The rationale behind the studies is that L2+ listening is a difficult skill and partly this is due to many learners having incomplete English phonology development. This is particularly the case in the author's context, English for Academic Purposes (EAP) for English Medium Instruction (EMI) in Japanese higher education. Many L1 Japanese, and other L1 Asian language learners of English, experience problems due to the larger vowel inventory of English. Such a difference in vowel inventory means that vowels are likely categorised as L1 equivalents, which may cause difficulties when two different English vowels are categorised as examples of the same L1 vowel.

An overview is given of the different models that inform the theoretical scaffold of the dissertation. The Perceptual assimilation Model (PAM; Best 1995) and PAM-L2 (Best & Tyler, 2007) are discussed, with their relevance for L2+ learners of English, as well as being a model which assumes all language categorisation occurs within the same template, i.e. all language categories are stored centrally and used across all languages. The Speech Learning Model (SLM; Flege, 1995) and its revision, SLM-r (Flege & Bohn, 2021) is

discussed as an example of what occurs with regard to high achieving learners; additionally, the SLM and SLM-r are of interest because they are models of phonetic rather than phonemic based learning. A further point of interest is that the SLM claims that first and second languages affect one another over time, with phonetic categories of L1 and L2 blurring. The Second Language Linguistic Perception model (L2LP; Escudero, 2005) examines the L2 state as it develops over time with exposure to L2 learning activities, ultimately culminating in an end state, and being able to be compared to the optimal state of language learning the L2(+). A further model, the Natural Referent Vowel framework (NRV; Polka and Bohn, 2011) explains why some vowels, in particular vowels with extremes of openness, closedness, and tongue protrusion/extrusion, are more readily learned than others.

A description of the methodologies of the studies used in producing the included articles is given: essentially following a pretest-post-test structure, although one study used a pretest-post-test-delayed post-test structure. As mentioned above, Bayesian analysis methods were used to analyse the data obtained. The reason for Bayesian methods, as opposed to more commonly used frequentist statistical analyses, is that Bayesian methods are more suited to the smaller sample sizes that are common in applied linguistics, and particularly classroom-based research. It is noted that the nature of all of this research is exploratory, and the case is made that this should be the default state of applied linguistics research, particularly classroom-based research, and that null hypothesis statistical testing is an ill-advised process with novel research (see also Trafimow, 2024).

General findings are that the use of Global Englishes as opposed to prestige varieties has no effect on phonology acquisition (Jones & Blume, 2022) or listening difficulty (Jones, 2024d). Additionally, preliminary evidence is found that suggests a link between visual salience and perceptual vowel acquisition (Jones, 2024c). However, the Bayes Factors in

the statistical analysis do not suggest strong evidence of a connection, therefore it is suggested that further research is required. Feedback timing is found to have a role in acquisition of L2+ vowels (Jones, 2024e), with delayed feedback being more beneficial for learning. Furthermore, a coincidental finding across the studies is that crammed learning appears to affect vowel acquisition, and has an interaction with learning of all target vowels and certain vowels depending upon the learning interventions.

The findings above are used to construct and/or support theories, such as a Bayesian model of phonology learning, where prior input is used as a reference when L2+ listening takes place. Such prior input is constantly revised, and thus parsing and listening skills develop. The dissertation concludes with the contributions to research and language teaching pedagogy and avenues for further research based upon the findings within the dissertation, some of which the author is undertaking as the dissertation goes to press.

Table of Contents

List of abbreviations

BF - Bayes Factor

BFA - Bayes Factor Analysis

BYOD - Bring Your Own Device

CALL- Computer Assisted Language Learning

EIL - English as an International Language

ELF - English as a Lingua Franca

EMI - English Medium Instruction

GELT - Global Englishes Language Teaching

GEs - Global Englishes

HVPT - High Variation Phonetic Training

L1 - First language

L2 - Second language

L2+ - Second or later additional language

L3 - Third language

L2LP - Second Language Linguistic Perception model

NRV - Natural Referent Vowel framework

OT - Optimality Theory

PAM - Perceptual Assimilation Model

PAM-L2 - Perceptual Assimilation Model for Second Language

PT - Processability Theory

PVs - Prestige varieties

SLA - Second language acquisition

SLM - Speech learning Model

SLM-r - Speech learning Model (revised)

WEs - World Englishes

List of figures

Chapter 1: Introduction and rationale

This dissertation discusses four journal articles and one conference paper related to three studies focusing on the perceptual acquisition of L2+ vowels by L1 users of Japanese studying English skills courses at an English medium instruction (EMI) university department. The dissertation project is important due to the situation of EMI in Japan: a growing number of students in the country are enrolling in EMI programmes (Ruegg, 2024). Additionally, EMI is not only becoming more widespread in Japan but also around the world. However, particularly in the Japanese context, students' skills base in English can be skewed toward reading comprehension at the expense of other skills. Furthermore, beyond Japan, listening has been reported as a weakness among learners of English for Academic Purposes (EAP) (Field, 2011). Incomplete phonology acquisition of the English vowel inventory is likely to play a part in such weakness because it is likely to lead to incorrect parsing of words. If this is the case, understanding the extent to which interventions in the studies detailed in the dissertation are effective in facilitating perceptual vowel acquisition can enable more effective teaching and learning. Ruegg (2024) notes that many students in Japanese EMI contexts enrol with the intention of improving their English. As such, many EMI departments in Japan provide English skills courses, which are necessary because L2 immersion does not necessarily result in improvements for all learners with the same exposure to the target language, affordances or opportunities (DeLuca et al., 2019). If more effective ways to facilitate phonology acquisition can be found, this aids not only the EMI sector but all language learning endeavours. That is, while the project is situated within a Japanese university EMI department's English skills curriculum, the findings are more likely to be transferrable to other contexts than not.

As a doctoral dissertation, the articles included are expected to be novel and to create new knowledge in the field of study. Arguably this is fulfilled by using the classroom context as the setting for second language acquisition (SLA) phonology research, and also in the case of Jones and Blume (2022), the remote, asynchronous context. The new knowledge gained is that the belief about 'native speaker' input being more advantageous for L2+ learning is debunked. Furthermore, the advantage of visual salience for phonology instruction is given credence with ecologically valid findings that take the benefits found in laboratory studies to the classroom. Similarly, the possibly counterintuitive finding in Jones (2024e) that delayed feedback is more beneficial for phonology acquisition should provide fuel for the critique of and improvements to computer-assisted language learning (CALL) materials development. While the sample sizes are small, steps were taken to make the findings

as robust as possible by using Bayesian statistical analysis methods, while also ensuring that the strength of what are initial, exploratory findings are not overstated.

Throughout the project, there are a set of assumptions that I make about the nature of L2+ acquisition:

- a) The first is that connectivism rather than generativism provides a more convincing theory of how languages are acquired, in that distributed learning based on frequency and salience of input in additional languages sufficiently explains the process in a simple way (Ellis, 2002), based upon constant language interactions in our surroundings from birth, as opposed to the leap of faith required to believe that language is simply innate (Dąbrowska, 2015).

- b) An additional assumption is that languages and their constituent parts can be interweaved and integrated within the different systems — that is, L1 features transfer to L2+ and vice versa (pace Flege, 1995; Flege & Bohn, 2021) and that L2 features transfer to L3+ and, again, vice versa (pace Gut et al., 2023).

- c) Furthermore, Pienemann's (1998) Processability Theory (PT) can, in my view, be applied not only to lexicogrammar and syntax but also to phonology and its acquisition in L2+. This is commensurate with a model of hypothesis testing posited by Faerch and Kasper (1980). The key way in which PT applies to phonology acquisition is that learners cannot process language items that they have not yet acquired, and equally, what has been acquired can be readily processed. For learners this means that exposure to salient examples of target phonological items needs to be provided in order for phonology to be acquired and, in turn, more accurate listening to be developed.

- d) Finally, I am using the term L2+ after Bassetti (2023), because global mobility is a fact of life for may of us, particularly those of us living in large mercantile cities. It is therefore difficult to find a truly monolingual person who has little to no experience of additional languages.

These assumptions form the backbone of my philosophy of language teaching, and therefore, when reading the articles which are the focus of this dissertation, these assumptions should be considered, even when they are not explicitly mentioned in the articles themselves due to considerations of journal word count limits and scope.

Overall Rationale

For hearing learners of non-signed, spoken languages, listening is an essential skill to acquire, as well as a modality in which one receives linguistic input. However, listening has been described as the most difficult of the four language skills to learn (Martínez-Flor & Usó-Juan, 2006). Whether this is actually true or not is difficult to ascertain; while it is the case that some L2+ language users perceive their listening skills to be the strongest of the four skills (Jones, 2024e), listening difficulties are not always noticed by learners (Nakatsuhara, 2011).

Nevertheless, due to the importance of listening in the language learning process, it is useful for language teachers and learners to understand how it can be better instructed. In my own experience as an English language teacher in to mainly L1 Japan learners for most of the last twenty years, it appears that many learners have problems in acquiring the ability to discriminate the English vowels /æ/, /ʌ/, /ɜː/, and /ɔː/. The misperception of these vowels can cause problems in vocabulary acquisition and, in turn, listening comprehension. This begs the question: how can the perception of these vowels be more effectively acquired.

The root of the problems appears to be related to perception and therefore attention to the acoustic content (i.e. phonetics), in relation to the articulatory gestures produced and intended as representations of items in the auditory language system (phonology). However, there may be more than simply phonetic and phonological factors involved in the acquisition of vowels.

Listening instruction takes place in multimodal environments, and this multimodality in the environment can influence the learning process. Visuals information is perceived, as actions coinciding with sound, articulatory gestures in speech, orthographic information (written text) in the environment and in learning materials. Considering these factors is undoubtedly important yet also rather difficult. With regard to Hayes-Harb and Barrios (2021) and Bassetti (2023), incongruent orthography used with auditory materials may hinder L2 phonological acquisition. One reason for this hindrance may be increased cognitive load (Sweller, 2008) due to competing and incompatible information as it adds strain to the working memory system (Baddeley, 1992; Baddeley & Hitch, 1974). Certainly, as Mousa-Inaty et al. (2012) found, there is a positive effect with the use of orthography for language acquisition prior to listening. Learners can choose whether to spend additional time on a language item or skip ahead in the text depending on perceived difficulty or affective factors. However, is this really an advantage of orthographic input or is it a disadvantage of the design of system or instructor-controlled media in listening tasks?

The primacy of orthographic literacy instruction due to progress in learning being more easily observed than aural instruction may mean that learners have already adopted strategies for dealing with L2+ orthographic input, but they may not have adopted strategies for aural input.

Conversely, use of visually salient input may be useful for language acquisition. The level of salience required for successful listening or acquisition is simply that which attracts sufficient attention to phonetic (acoustic) features. In particular, visual information related to articulatory gestures (i.e. movements of the vocal tract) should be able to facilitate acquisition of vowels, based on neuroscientific research (Glanz et al., 2018; Hertrich et al., 2020). Sueyoshi & Hardison (2005) and Hardison (2017) obtained results illustrating that gated video of speakers' mouths facilitated greater gains in acquisition of English stop phonemes by L1 Japanese and Korean learners of English when compared to audio-only input. Perceivably different articulatory gestures between English and French /u/ vowels have also been successfully trained under laboratory conditions for French L1 and English L1 participants respectively (Masapollo et al., 2017). Lee and Mayer (2015) found that contextual video added to an audio lecture improved comprehension for L1 Korean learners of English. However, Cross (2011) found that L1 Japanese learners of English paid little attention to "talking head" videos, which illustrates a difficulty in persuading learners to do what is in their pedagogic best interests.

Listening difficulties for L1 Japanese learners of English

Listening in L2+ is difficult for many language learners, especially when there are major systematic differences between L1 and L2+. Japanese and English have several differences, particularly syntactic structure, the use/omission of pronouns due to different expectations regarding context, and also the difference in size of the phonological inventories. English has not only a larger number of phonemes than Japanese, but these extra phonemes are mainly vowels. This difference is likely to be one of the contributing factors toward L1 Japanese learners' difficulties with English listening.

When L2+ learners encounter a stream of speech and they wish to attend to it, they will attempt to decode it using their phonological knowledge. If L2+ categories are not sufficiently integrated into the perceptual phonological template, listening will

be carried out largely using L1 phonological categories (pace Best & Tyler, 2007). In doing so, lexical items are likely to be subject to one the following conditions:

- parsed (i.e. all phonemes are the same across L1 and L2+);
- parsed with delay (context is high and learners may be familiar with the item in orthographic form);
- misparsed (wrongly identified, e.g. 'cart' /kaːt/ as 'cat' /kæt/); or
- unparsed (no parsing can take place due to low context and/or incomplete L2+ phonology acquisition).

Bonk (2000) found that the Japanese learners of English that participated in his study on lexical identification could not identify words that they were familiar with from reading when asked to listen at the utterance level. Furthermore, Joyce (2013) identified word recognition as a factor in listening comprehension, which is supported by the work of McClean et al. (2015) and Carney (2021). Indeed Carney suggests that learner misperceptions based upon errors in phoneme discrimination or word segmentation may continue upon repeated listening even when semantically illogical. Although it would seem logical that if words are known to the listener through reading, then orthography may provide assistance, but this may not actually be the case.

While studies have shown that orthography can have an effect on listening/phonology (Barrios & Hayes-Harb, 2020; Basetti, 2023; Hayes-Harb and Barrios, 2021; Showalter & Hayes-Harb, 2013; Wolfgramm, Suter & Göksel, 2016), the influence can be either catalytic or obstructive depending upon the L1-L2+ language pairings, familiarity with orthography, the transparency of grapheme-phoneme correspondences (GPCs), and the phonological items to be acquired. Japanese orthography is more or less a one-to-one mapping of grahemes to mora consisting of either monophthong vowel, consonant + monophthong, or a single consonant representing /n/ used at the coda of a syllable. However, English has a particularly opaque orthographic system, with many-to-many mappings, even with regard to monophthong vowels, and therefore orthographic learning can interfere with phonology learning (Bassetti, 2023; Mathieu, 2016). As such, despite familiarity with English orthography, L1 Japan learners of L2+ English may be better served by alternative methods to raise attention to the target phonological items that they may be unable to perceive or discriminate between in media or from interlocutor speech.

The articles

The articles and conference paper forming the basis of the dissertation use a series of interventions to investigate the effects on L2+ vowel acquisition for the facilitation of more effective listening. The articles and conference paper can be found within this dissertation in appendices 1-5 (pp. 99-193). The interventions are: the use of Global Englishes (GEs; Galloway, 2013; Rose & Galloway, 2018) versus prestige varieties (PVs) of English, the use of visual salience to catalyse perceptual phonology acquisition, and feedback timing with regard to vowel identification.

The reason for choosing GEs is one of practicality: the participants were mainly L1 Japanese students in an English-medium instruction (EMI) university department in Japan, with 30% international students making up the student body. In addition, many of the faculty (the author included) are non-Japanese (although speaking standardized varieties of English, arguably PVs, in the English language teaching section), and teaching assistants in the department's self-access learning centre have come from outer-circle or expanding-circle countries according to Kachru's (1988) classification. This being the case, the domestic students need to be able to listen accurately in order to work effectively with international classmates and to understand the speech of the diverse faculty. Munro (2021) stated explicitly that (non)native-speaker status in not a condition that relates to the ability to communicate effectively. By conducting a study using GEs and PVs, the basis for the predominant use of PVs can be evaluated for pedagogical soundness in comparison to GEs. To conduct the investigation, a quasi-experimental study was carried out, with a pretest-post-test structure. Tests consisted of 32 vowel identification task items. Learners in the intervention group were provided with input from GEs speakers, and a control group of learners were provided with input from PVs speakers. A comparison could then be drawn between the groups in order to evaluate whether there was any meaningful difference in acquisition of English vowels.

The decision to investigate the effects of visual salience upon perceptual vowel acquisition was based upon the strength of evidence in the literature regarding consonants. Hardison's (2007; 2018) work in a laboratory setting with gated video, that is video played and stopped in stages, showed that making consonants salient with audiovisual input facilitated greater gains in consonant acquisition than audio-only input. In transferring the laboratory study rationale to the classroom, the methodology was untenable due to the artifice involved with only the speaker's mouth visible on screen. Additionally, the use of gated video is somewhat difficult and

not a naturalistic way of parsing speech, therefore isolated consonant-vowel-consonant words (CVC words, e.g. 'cat' /kæt/) were used prior to having learners listen in an utterance context. In the study which forms the basis of Jones (2024c), learners received identical speech as input, but with the intervention group receiving the information in the audiovisual modality (as embedded videos in a Moodle environment) and the control group receiving the information as audio only (as embedded mp3 recordings, again in Moodle). The target vowel learning gains were then compared using pretest-post-test, and delayed-post-test gains as a comparison, with a battery of 32 vowel identification items used as the test.

In the final study, learners in both groups were provided with identical input from edited TED videos, with feedback provided either immediately, or delayed until after a series of questions forming a Moodle lesson were completed. The learners' gains in perceptual vowel acquisition were again measured using pretest-post-test gains on a vowel identification task battery.

The focus of this dissertation is upon the following questions related to listening and phonology acquisition:

RQ1. Does a speaker's variety of English have an effect upon (a) vowel acquisition, and (b) ease of listening for L2+ learners of English?

RQ2. Does input modality affect L2+ perceptual vowel acquisition?

RQ3. Does feedback timing affect L2+ perceptual vowel acquisition?

Chapter 2: Literature Review

Phonology

Phonology is usually thought of as two strands: (1) segmental phonology, or the different individual abstracted sounds of speech, usually referred to as phonemes; and (2) suprasegmental phonology, or the patterns of tone, volume, and other shifts of character that the phonemes undergo as they are realised. This dissertation is mainly concerned with segmental phonology, and vowels in particular. Whether phonemes are the correct unit to focus on is not straightforward; Vihman and Croft (2007) posit that children learn "one or a small number of phonological templates" (p. 686) based upon words and add to them over time as they are exposed to and attend to language. Note that their theory is based upon L1 and L2+. Conversely, Strange, Levy and Law (2009) state that learners are capable of working at the phonetic level, i.e. that learners can work with the exact acoustic information without categorising the sounds as part of the L2. However, Mitterer and colleagues (Mitterer et al., 2013; Mitterer et al., 2018) state that learners should be able to understand allophones, or variables that are realised in complementary distribution, either through categorising them as similar to L1 speech or otherwise (c.f. Best, 1995; Best & Tyler, 2007). Furthermore, Mitterer et al. (2018) have results to support allophones being processed at the pre-lexical stage as opposed to phonemes. That allophones can be perceived by L1 users of the language is one thing; it is quite another to assume that L2+ users process at anything other than a level beyond the L1 phonological template and the current stage of L2+ phonological development. However, phonemes and phoneme categorisation are frequently used in both research and teaching. Whether this pre-lexical processing applies to language teaching is debatable.

It should be noted that phonology is a key part of the working memory model (Baddeley, 1992; Baddeley & Hitch, 1974). The model consists of a phonological loop, a visuo-spatial sketchpad, and a central executive, which controls function and decision making. As such, working memory also plays a key part in listening, particularly in L2+. For L1 listening , which can be challenging itself, sound is perceived and then held in working memory during parsing, and arguably during response. In L2+, there may be a need to process not only unfamiliar phonology, but also phonetic items that have not yet been categorised as phonemes. This processing can add strain to the working memory and therefore is likely to result in a breakdown of the parsing process.

The main models of phonology acquisition that provide theoretical underpinning to this dissertation are the Perceptual Assimilation Model (PAM; Best, 1995) and the offshoot models PAM-L2 (Best & Tyler, 2007; Tyler, 2019) and the Vocab model (Bundgaard Nielsen et al., 2011, 2012); the Speech Learning Model (SLM; Flege, 1995) and the revised model (SLM-r; Flege & Bohn, 2021); and the Second Language Linguistic Perception model (L2LP; Escudero, 2005). A further, more specialised model related to vowel acquisition is the Natural Referent Vowel framework NRV; Polka & Bohn, 2011).

In the PAM (Best, 1995), a direct realist view of phonology is used, which means that a complex interface and reconstructions of perception and production are eschewed in favour of a direct observation of speech. This direct observation then provides the basis of how phonemes are categorised by the listener. The PAM posits different types of categorisation of phonetic contrasts which are:

1. *Two Category Assimilation*: each segment assigned to different native category.

2. *Category-Goodness Difference*: Each segment assigned to same native category but differ in deviance from ideal. Discrimination from good to poor depending upon distance.

3. *Uncategorized versus Categorized*: One segment categorized in a native category, one not. Discrimination good.

4. *Nonassimilable*: Both non-native sounds categorized as non-native. Discrimination good. (Best, 1995)

For most teachers of language, the categorisations most of interest are 1 and 2, because speech sounds that are so unlike speech are likely to be rare, or in the case of click languages, learners are likely to be instructed that the language contains clicks and therefore will attempt to categorise those sounds.

The way that this categorisation works, according to Best (1995), is that listeners are used to ignoring inconsequential variations in L1 speech, and such attunement to the L1 creates difficulties in L2+ categorisation. According to the PAM-

L2 (Best & Tyler, 2007), naïve listeners can only differentiate between their own L1 phonemes and deviations from them. In contrast, "the phonological level is central to the perception of L2 speech by [second-language] learners, who are developing an L2 (or interlanguage) system, in a way that it cannot be for L2-naïve listeners perceiving unfamiliar nonnative speech" (p. 23, square brackets mine; parentheses in original), and as this develops, more phonemes can be discriminated depending upon the state of the learning development. Best & Tyler also state that the formation of L2 categories is not required for L2, instead perceiving "higher order invariants in speech stimuli" (p. 25) such as formant values in the case of vowels. However, formants are related to the physical arrangement of the articulators in the vocal tract, i.e. articulatory gestures, which Best and Tyler also say can be perceived.

A step beyond this increased perception of L2 phonemes may occur due to "contrast assimilation type changes as a result of new category acquisition (e.g. a category goodness assimilation becomes a two-category assimilation)" (Tyler, 2019. p. 611). This is a progression from Best and Tyler (2007), which stated that the mental representations were unnecessary for perception. Obviously, perception occurs before (and as a direct prerequisite for) categorisation, but as further categorisation occurs, newer L2+ percepts can potentially be mapped in the phonological system. The above factors also interact with word learning and the PAM-L2 was extended to the Vocab model (Bundgaard-Nielsen et al., 2011, 2012), which states that as vocabulary is learned, the categorisation of L2 phones is also fine-tuned. This provides a basis for what happens with L2+ learners beyond the early phase of their learning.

When new sounds are perceived, they can be added to the phonological template. The phonological template, in PAM, is used for all languages in the user's repertoire (Best, 1995). In this way it is somewhat different to the SLM (Flege, 1995) which is discussed below. However, Best and Tyler (2007) elaborate on a popular misconception about the similarities of the PAM and SLM:

> PAM was developed specifically to explain nonnative speech perception by naïve listeners, whereas SLM was designed to address production/perception of L2 speech by L2 learners. Neither model was developed to address both situations, although each has been cited as though it had been.
>
> (pp. 21-2.)

Furthermore, regarding the weaknesses of the PAM-L2, and its focus on second language (i.e. not foreign language learners outside the target community), Tyler (2019) comments that "The principles of PAM-L2 were illustrated using the idealised situation of a learner previously naïve to the L2 in an immersion environment. Such a learning situation may be rare in the modern age, especially when the target language is English" (pp. 615-6).

One of the major differences between the PAM and SLM is that the PAM operates on a phonemic level whereas the SLM operates on a phonetic level. That is "as learners gain experience in the L2, they gradually discern the phonetic difference between certain L2 sounds and the closest L1 sound(s). When this happens, a phonetic category representation may be established for the new Lw sound that is *independent of representations established previously for L1 sounds*" (Flege, 1995, p. 263, my emphasis). In other words, the L2 category is separate from all L1 sounds, which is akso different to the PAM, which states that all speech categories are located on the same template. Guion et al. (2000) state that while the PAM framework can be extended "the naturalistic acquisition of English as an L2" (p. 2723), the SLM could not be extended to account for the acquisition of L2 sounds by inexperienced learners. This remains the case with SLM-r (Flege & Bohn, 2021), which still focuses upon primarily migrant/long-term L2 users rather than the relatively inexperienced learners/users focused upon by the PAM and PAM-L2.

One proposition made in the SLM is that learning can be predicted by considering the distance between the target L2 phone and nearest L1 phone. Over time, L1 and L2 affect one another and similar phonemes from L1 and L2 blur together (Flege, 1995). However, the SLM-r (Flege & Bohn, 2021) states that speech perception depends upon adapting phonetic categories in real time according to prior input, "even the recent past" (p. 15), based upon statistical distributions of input. However, L2 and L1 are not learned in the same way, due to L2 sounds initially linking to sounds in L1 and possibly also being interfered with and blocked by those L1 categories. They additionally propose a three-stage process for speech sound acquisition: first, discerning a phonetic difference between an L2 sound and the closest L1 sound; second, an essential equivalence between those sounds is necessary; finally, the link between the equivalents is broken and the L1 and L2 categories are perceived separately. However, the break between L1 and L2 categories is a long-term process and as such, it is worth remembering that "The SLM-r shares the view of other theoretical models (e.g. Best & Tyler, 2007) that L2 speech learning is profoundly shaped by perceptual biases induced by the L1 phonetic system" (Flege & Bohn, 2021, p. 24).

In short, while the PAM and PAM-L2 model the phonological learning that relatively inexperienced language learners undergo, the SLM and SLM-r model the phonetic learning experienced by long-term users of the L2+ in a community where that language is used as the main language of communication, although individuals may spend large amounts of time with other L2+ users of the language, which can also influence the process of their language acquisition.

Initially outlined in Escudero (2005), the L2LP model posits that L2 is perceived by the listener, initially in L1 mode. Although earlier versions of L2LP made use of Optimality Theory (OT; Prince & Smolensky, 2002), which states that speakers/listeners optimise their choices of phone dependent upon what is most likely to allow successful decoding. Yazawa et al. (2020) used Stochastic OT, which provides allowance for probabilistic variations unlike traditional OT. The model was later revised in Van Leussen and Escudero (2015). The five components of the L2LP are:

1. optimal L1 perception and optimal target L2 perception
2. the L2 initial state
3. the L2 learning task
4. L2 development
5. the L2 end state

Optimal perception is defined as "the optimal way to perceive the sounds of a language is by making categorization decisions that lead to maximum-likelihood behaviour" of understanding the phones in question (Escudero, 2005, p.52). The initial state is, of course, the state that learners are in when they are of interest to the researcher and the L2 learning task is the task of interest to the researcher, which is also self-explanatory. L2 development is a state which is the result of the initial state + the L2 learning task, with the L2 end state essentially the final result of L2 development(s). It can be somewhat difficult to understand the model particularly the difference between L2 development and the L2 end state. However, not all components

necessarily progress and develop. According to Escudero and Boersma (2004) “L2 listeners

may reach optimal perception or may manifest suboptimal optimization strategies that are

specific to L2 acquisition” (p. 555).

Regarding phonological categorisation, the model allows for four types of mapping:

1. Auditory-to-auditory mapping constraints
2. One-dimensional auditory-to-feature constraints
3. Multidimensional auditory-to-feature constraints
4. Multidimensional auditory-to-segment constraints

These multiple types of mappings are important because “targets of perception need to be abstract enough to allow for the integration of multiple cues and not just for the mapping of one auditory continuum onto a perceived auditory category along that same continuum” (Escudero, 2005, p.45) because adults attend to many different types of cue. Cues come from distributed learning (Williams & Escudero, 2014) which also allows the model to take into account L3 learning. Similar to the SLM(-r) (Flege, 1995; Flege & Bohn, 2021), there is an assumption of different “perception grammars” (Williams & Escudero, 2014, p. 10), which is different for PAM, which posits a shared phonological map for all languages.

Further contrasts between the models are that “SLM would explain that, for some

learners, equivalence classification resulted in persistent use of duration, whereas for others,

noticing phonetic differences between the L1 and L2 categories led to new category

formation”; however “From the perspective of PAM(-L2), some learners may have equated

the L1 and L2 contrasts both phonetically and phonologically, whereas others may have

equated the contrasts only phonologically while being aware of the phonetic dissimilarities"

(p. 575).

Common points of the models discussed are that phones are likely to be 'mapped' to L1 templates, either the same (Best, 1995; Tyler & Best, 2007) or a copy of it (Escudero, 2005; Flege, 1995; Flege & Bohn, 2021) and over time this changes to an L2 template that is somewhat less dependent upon L1. What remains unclear is just which Lw phones or phonemes are likely to be a part of a learner's interlanguage phonology (pace Selinker, 1972) at a given stage in their language learning. In other words, we lack a L2 phonological acquisition sequence for any L1-L2 pairs of languages, including the context with which I am concerned primarily, L1 Japanese-L2+ English.

Adult listeners use several types of auditory dimensions as cues, constraining perceived auditory information through mappings to audio percepts, phonetic features and to phonemic categories (Escudero, 2005, p. 45-7). According to Escudero's (2005) Language Perception (LP) model, the output of perception is phonological and not yet related to lexical items, and thus still abstract (p. 51). However, for L2 perception "one of the L2LP's basic claims is that a separation of perceptual mappings from sound representations leads to an adequate comparison of the perception systems involved" (p. 86); that is, when comparing what is heard, these abstract mappings allow listeners to compare the speech heard to prior experience based on auditory percepts, phonetic and phonemic features as mentioned above. This theory matches the phenomenon of novice L2+ being able to mimic L2+ phonological items, yet due to insufficient exposure being unable to adequately acquire them for parsing after the initial exposure.

The Natural Referent Vowel framework (NRV; Polka & Bohn, 2011) is different from the models above but arguably at least as important for this dissertation, not least because the focus is on vowels. The main idea of the NRV is that the most peripheral vowels in the vowel space work as referents to which other vowels are categorised. As such, "an initial bias favoring vowels falling close to the corners of the vowel space is clearly advantageous no matter what language is to be acquired" (p. 473). While conceived as a theory of infant L1 acquisition, the authors state that:

> [this] bias may also reappear when perceivers are mapping out a new vowel system to learn a second language, possibly when new L2 vowel categories fall in regions of the vowel space that are not firmly committed to the L1 (p. 474).

Although the bias exists, it can be overridden by mature perceivers.

In studies described in Polka et al. (2019), video stimuli and photographs were used to stimulate L2 learning of vowels. Video was attended to much more when perceiving peripheral vowels. Photographs of speakers' faces did not result in a perceived asymmetry in vowel perception, suggesting that photographs were insufficiently salient. They also claim that this is evidence that the NRV is based on phonetics (including articulation), and not simply acoustic data.

Phonology is acquired for both receptive and productive reasons, with receptive phonology acquired prior to productive phonology (Pennington & Rogerson-Revell, 2019, p. 72). Although learners may be able to mimic examples of speech sounds (Baese Berk & Samuel, 2016), they are unable to produce the sounds without and example of the phonetic details as a cue. Rather, learners require a phonological inventory sufficient to understand the phonological units that a speaker is using (Strange, 2011) in order to parse the units at a lexicogrammatical level. Due to phonology being acquired for reception prior to production, care must be taken when using productive capabilities as a proxy for perceptual capabilities. Production may be used as a proxy but it does not provide all the details of a learner's abilities, only what they are able to produce. However, Kartushina and Frauenfelder (2014) found that vowel perceptions had little relation to L2 pronunciation of some similar vowels among Spanish school learners of French, therefore caution is advised.

Phonological development is important for teachers because learners' problems cannot be solved if there is no inventory available to describe their problems and successes. With an idea of learners' base phonological inventories, materials can be selected for efficient learning. Although there are studies on the order of of phoneme acquisition for children speaking English as a first language (Salus & Salus, 1974) there are no such studies available for learners of English as a second or foreign language. The reasons for this absence are practical: it is too difficult to control for exposure to other languages; different languages pairs have different

overlaps. Due to the absence of a known order of acquisition, it is therefore prudent to compare learners' L1 phonology with English in order to prioritise the phonemes that are unavailable for direct transfer from the L1.

Vowel perceptions

There are clear patterns in the psycholinguistic research regarding perceptual discrimination of vowels. Work initially conducted in Barcelona (Bosch, Costa, & Sebastián-Gallés, 2000; Pallier, Bosch, & Sebastián-Gallés, 1997; Pallier, Colomé, & Sebastián-Gallés, 2001; Sebastián-Gallés & Soto-Faraco, 1999) and supported by results in Flege & MacKay (2004) found that childhood L2 learning does not necessarily confer a "nativelike perception of L2 vowels" (Flege & MacKay, 2004, p. 26) yet learning L2 after the L1 is established does not necessarily result in measurable differences between L1 speakers and early L2 learners. Gilichinskaya and Strange (2010) found that Russian L1 listeners generally categorized American English vowels in correspondence to the nearest Russian equivalent. Levy and Strange (2008) found that experienced L1 American English

users of L2 French still made errors with vowel perception when vowels were proximal to one another in the vowel space (i.e. mouth and tongue position), and also that there were differences in probability of error dependent upon other phonemes they were adjacent to in connected speech. Such errors were more likely among inexperienced language users than experienced language users. It therefore appears that L2+ vowel perception requires instructions, even for experienced learners, although increasing experience with English listening may mitigate errors in perception.

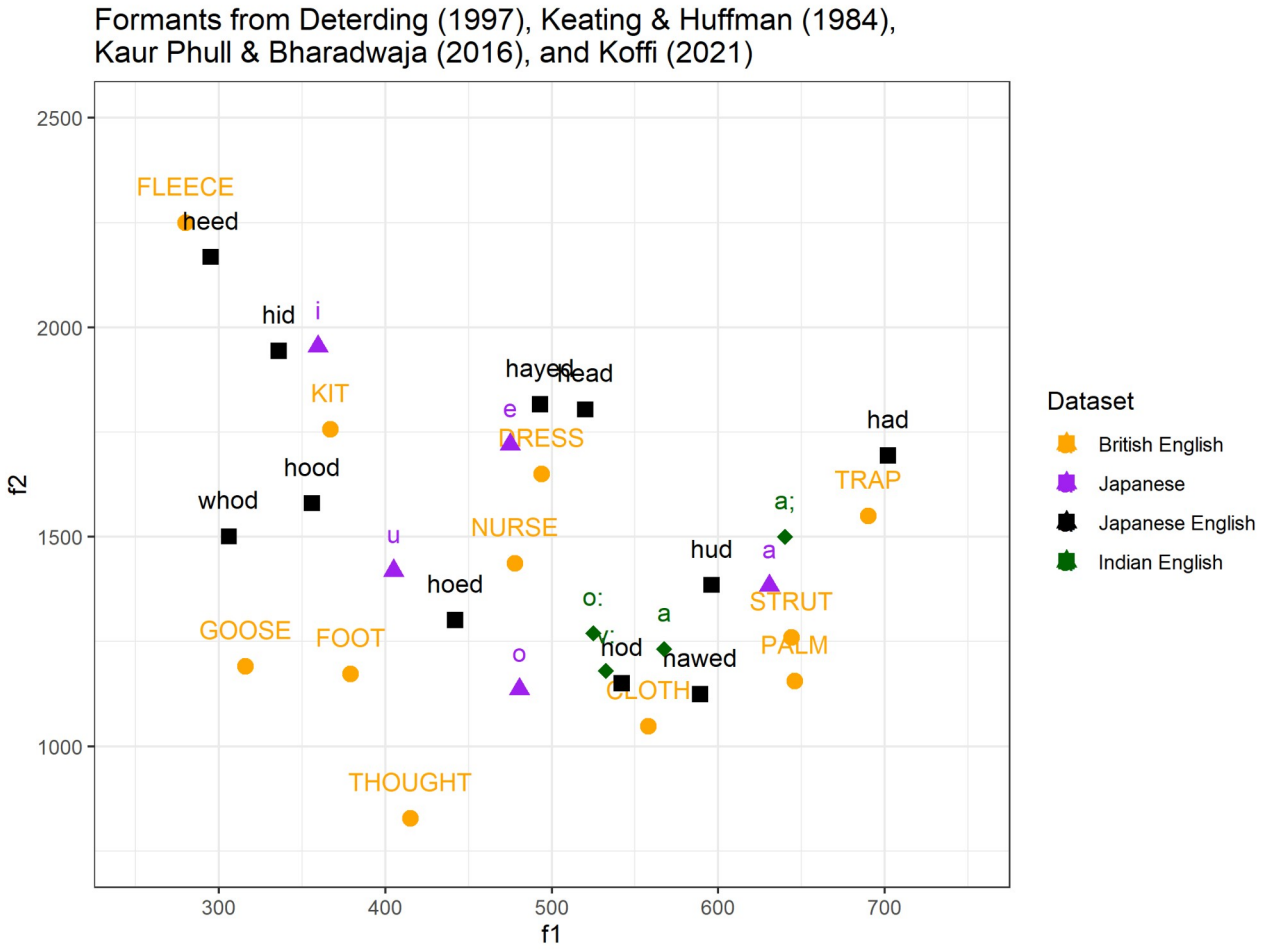
One interesting pattern in vowel perception patterns comes from Polka and Bohn's (2011) Natural Referent Vowel (NRV) model, which states that movement from a central vowel (mid-open, tongue in mid position) toward a more peripheral vowel (a vowel either more open/closed, with tongue more front/back). For example, movement from English /ʌ/ to /ɑ/ is easier to discriminate than the opposite movement. In other words, proximal vowels must be contrasted, and contrasts must be given in both directions.

Additionally, Nishi & Kewley-Port (2008) found that Korean L1 learners identified more American English vowels correctly when provided with a wider

inventory than when provided with a limited set. Thus, while a limited number of phonological items may be targeted for instruction, it is pedagogically prudent to avoid artificially restricting the number of items presented in natural speech or in recorded materials. However, if orthography is to be avoided, a large amount of on-screen real estate is necessary for picture or symbol representations, and therefore pedagogical methods require careful consideration, particularly in CALL and/or TELL contexts, in order to avoid limiting inventories of vowels overall that learners are exposed to. However, phoneme choices in identification tasks may need to be more limited than the range of stimuli overall for practical reasons.

Vowels can be difficult for Japanese learners of L2+ English to perceive. Such learners have the consonants of English available to transfer from L1, with the exception of /ð/, /θ/, /ʒ/, /ŋ/, and the /l-r/ contrast, which has received a large amount of attention (e.g. Lively, 1993; Lively et al., 1994; Lotto et al., 2004; Sheldon & Strange, 1982; Shinohara, 2016). Additionally, those phonemes that are not subject to an easy transfer from L1, other than /l-r/, are not particularly difficult for L1 Japanese learners to make reasonable guesses about due to a combination of relatively low functional load (how they differentiate lexical items), relatively low frequency in speech, and in the case of /ŋ/, frequently collocating with /g/ so much so that learners verbs in the continuous mood, as in the current standard British pronunciation described by Lindsey (2019). However, with small shifts in height or backness, vowels change, and the English vowel inventory is larger than that of Japanese (c.f. Deterding, 1997; Kaur Phull & Bharadwaja Kumar, 2016; Keating & Huffman, 1984; Koffi, 2021). Note that these values are generally means of each from recordings. Additionally, as Flege and Bohn (2021) state, adult L2 learners are likely to be exposed to more varied L2 input, including more speakers of different varieties of the L2, than are L1 monolingual children working to establish a phonology of the same language.

Figure 1: Vowel first formant (f1) and second formant (f2) plot frequencies from Deterding, (1997); Kaur Phull and Bharadwaja Kumar (2016); Keating and Huffman (1984); and Koffi (2021).



As can be seen from Fig. 1, the Japan vowel space is narrower than that of English, with no vowels having as high or low first formant (f1, mouth openness) or second formant (f2, tongue front/backness) values as those of British English. However, Japanese English has similar tongue front/backness values as British English, but different mouth openness values. Notably, there are three Indian English vowels, /ɔː/, /ʌ/ and /a/ (noted as 'o:', 'v;' and 'a' respectively) clustered around

Japanese English /ɒ/ (hod). Such clustering, along with unfamiliarity with Indian English, potentially accounts for the anecdotal difficulty among L1 Japanese learners of English in discriminating Indian English speech sounds accurately. However, British English vowels also cluster around Japanese L1 phoneme values and therefore similar confusions are also likely for this variety, despite it being accorded more social prestige (Rubdy, 2015), and being more prevalent in English language teaching materials (Kiczkowiak, 2021). The reason that this comparative phonology is important is due to what has been criticised by Kubota (2015) as a focus upon the intelligibility of GEs speakers, when there is even considerable variation in 'native speaker' varieties and even among and within those speakers themselves. By shifting focus from speaker to listener, and teaching to widen phonological templates, intelligibility should be made easier as listeners can parse speech from a variety of speakers more easily.

Frequency-based accounts of acquisition (e.g. Ellis, 2002), connect greater learner familiarity of linguistic features to greater ease of discrimination and, in turn, acquisition. In extending this logic to language varieties, greater familiarity with a given variety should enable perceptual acquisition and render comprehension easier. Goh (1998) investigated the use of listening strategies by L1 Chinese learners of English, and found that novice listeners tended to use more bottom-up processing strategies than more accomplished listeners, even though the strategies did not necessarily result in accurate comprehension. That is, due to a lack of overall listening proficiency, the learners needed to make use of basic perception strategies to attempt to decode speech. However, phonological structures that learners may not be accustomed to hearing may take longer to acquire than those which are similar to L1 (Archibald, 2021). Therefore, if the target phonology is not developed sufficiently to parse phonemes effectively, listening for details and overall meaning are hindered. Furthermore, Galloway (2013) attested to student participants' reported unfamiliarity and difficulty in understanding British English despite being third and fourth-year English majors at university, and actually taught by a British English speaker. This being the case, instruction in varieties of English that are likely to be encountered by learners should be prioritised, particularly when they differ markedly from L1 inventories. It is therefore important that language teachers and learners attend to phonology because an ability to perceive speech and discriminate sounds accurately requires less strategy use due to an ability to decode and parse speech more effectively.

When learning vocabulary with little or no phonological input, learners are only creating an incomplete lexicon, and their receptive skills will be lacking because accurate phonological forms would not have been acquired. The point was made by

Bonk (2000) and again by Joyce (2013), that Japanese learners do not recognise the phonological forms of words they know in orthographic form, particularly in connected speech. While the focus in this dissertation is on vowel learning, my experience as a teacher is that phonology in general and vowels in particular are not focused on in students' learning, due to a persisting overall focus at the high school level on university entrance examinations (Kikuchi & Browne, 2009; Kobayashi, 2001; Yamanaka & Suzuki, 2020). This neglect of phonology then leads to ongoing problems with listening, particularly learners attempt to parse words from authentic speech.

Phonology teaching

Teachers are often faced with learners in their classrooms who have problems in listening to speech in their second language. Although it is not the whole reason that learners have difficulties in listening, incomplete L2 phonology acquisition can be a significant factor in this difficulty. According to survey research conducted with teachers worldwide (Jones, 2020), bottom-up listening skills and phonology are considered by a significant minority of teachers, but a minority nevertheless. Below, the literature related to phonology acquisition is reviewed and classroom interventions are suggested in the hope that they may bring about greater success in perceptual phonology acquisition for listening.

Work on consonant perception and acquisition appears to be abundant. In first language acquisition, Nazzi (2005) found that consonant contrasts make word learning easier than vowel contrast, regardless of vocalic distance of the vowels (i.e. tongue position, openness and rounding of the mouth). Wang and Chen (2020) found that phonetic distance can cause problems in perception; palatal and retroflex phonemes were frequently confused by L1 English learners of Mandarin Chinese. However, Sueyoshi and Hardison (2005) and Hardison (2003), found that stop consonant (/p/, /t/ and /k/) perception is improved in learners through gated video. This leads to a conclusion that phonology perception training requires further classroom based-research, in order to understand how phonology can be taught effectively.

According to Strange (2011) "perception is an activity with a purpose. In everyday communication, listeners attend to incoming speech signals with the goal of perceiving the linguistic message that the speaker intends" (p. 457). To this end, L2 listeners, likely have ideas and expectations about the semantic content of a speaker's

message in a given context. However, they are also likely to have preconceptions based upon their internalized interlanguage (Selinker, 1972), where their current state of perceptual phonology dictates perception of certain syllables in lexicogrammatical form. One point of interest made by Strange (2010) is that "Our own studies of L2 learners of AE [American English] have shown that vowel perception performance does not correlate significantly with either length of residence in the USA, or with self reports of L2 use in everyday life. However, vowel perception accuracy does correlate significantly with tests of English language proficiency" (p. 464.) Strange also states the need for more work on how to train learners on phonetic discrimination. For language teaching research in particular, further research on training would provide a stronger evidence base for teachers, administrators and course designers to draw upon andapply in context for specific communities of language learners. This challenge has been taken up in Spain in particular, with research into High Variation Phonetic Training (HVPT).

HVPT has been mooted as a method for improving learners' segmental phonology acquisition. Certainly Mora & Mora-Plaza (2019) found that Catalan/Spanish learners of L2+ English improved vowel contrasts after undertaking HVPT, with Cebrian et al. (2009) finding that perceptual abilities increased among HVPT learners. One weakness of HVPT is that it relies on uncommunicative laboratory-derived processes, particularly its use of non-word repetition and identification, and therefore may not be taken to with enthusiasm by learners. Additionally, whether HVPT can be used in tandem with visually salient joint input (i.e. still photography of the mouth during articulation or video of speech) in order to accelerate acquisition remains to be seen. Furthermore, whether HVPT can be introduced at any point in learning without disruption to acquisition is not wholly clear, given that Baese-Berk & Samuel (2016) found that repetition of stimuli to be learned by naïve Spanish learners of Basque disrupted Basque learning, although not in a HVPT context, . Clearly, bearing in mind the differences between the studies on HVPT and those of Baese-Berk & Samuel (2016), such as L1-L2 pair differences, caution must be shown in how one plans to provide phonological learning affordances.

A different method to facilitate phonological acquisition, which is prevalent in Japanese institutional ELT, is shadowing. Hamada (2016) makes claims for building bottom-up listening skills. Learners focus on the sounds a speaker makes and attempt to reproduce them as close to possible. The aim is that through ignoring meaning, learners must focus on phonetic and/or phonological phenomena. Despite Hamada's (2014) claims that shadowing facilitates acquisition, particularly after learning content through reading, the testing procedures used the TOEIC test, without

accounting for either the effects of vocabulary knowledge from reading or test-wise strategies among the university student sample, which reduces the validity of claims for effectiveness. As with any methods that claim to improve phonological awareness, more research with targeted phonological testing procedures is required to assess the validity of shadowing.

Perhaps surprisingly, one of the biggest impacts in phonology teaching that is rarely

transferred to practitioner-oriented literature is the use of reading and orthography with learners and its effects upon phonological discrimination. “Many L2 speakers are already literate and often experienced readers when they begin acquiring a second language, so may tend to pronounce L2 words based on L1 pronunciation as tied to L1 orthographic patterns, as an aspect of L1 transfer” (Pennington & Rogerson-Revell, 2019. p. 185). This may be a major factor not only in ESL contexts but also in EFL contexts. It may be assumed that in ESL contexts, learners are ‘immersed’ in the L2; however, in EFL contexts, i.e., where it is assumed that learners have less access to target language media, English is still pervasive. This is particularly true when considering the linguistic landscape (Backhaus, 2007; Landry & Bourhis, 1997), that is, the use of language in signs, announcements, logos and other media that are part of everyday life, particularly in urban areas. Even in locations where L1s do not use Latin orthography, learners are still extremely likely to be familiar with corporate branding from store fronts, etc. and associate the orthographic information with the typical pronunciation of the entity name in the L1. There is also a lack of research into whether loanwords from L2 in the L1 can negatively affect the acquisition of phonological rules. For example, in Japanese the loanword ‘sutorongu’ is a borrowing of the English ‘strong’, and Chinese ‘kafei’ is a loanword from French, but is a calque for coffee. However, Japanese uses a phonological syllabary for loanwords whereas Chinese makes use of ideograms which usually reflect the approximate phonology of the loanword. Whether loanword orthography has an ongoing is yet to be determined. However, in experimental studies orthographic effects upon phonology have been observed.

Hayes-Harb and Barrios (2021) reviewed the literature on orthographic effects on L2 phonology and found that benefits of orthography may only be found in orthographically congruent L1-L2 pairs, which is also supported by Han and Kim (2020). That is, where languages share orthographic representations (e.g. English and German to a certain extent), it may be possible to use the L1 orthography to assist in L2 phoneme learning. However, among incongruent pairs, such as English and Japanese, the reverse may be the case. For example, Sokolovic-Perovic, Bassetti and

Dillon (2019) found that orthographic geminate consonants (double-lettered consonants e.g. the /p/ as 'pp' in 'shopping') were lengthened by L1 Japanese learners of L2 English when provided with orthographic guidance for words. Mathieu (2016) found that Arabic orthography inhibited L2 Arabic phonology learning by L1 English speakers with no prior knowledge of Arabic. Other studies with L1 Japanese learners of English by Bonk (2000) and Joyce (2013) also found that their learners could comprehend words in isolation but could not comprehend them in natural spoken English. Both Bonk (2000) and Joyce (2013) ascribe these findings to unfamiliarity with aspects of connected speech in English. However, because orthographic interference has not been extensively investigated, it may not be ruled out that orthographic overreliance or overfamiliarity has brought about this observed phenomenon. Interference from orthography in phonological acquisition has been documented by Showalter and Hayes-Harb (2015) with regard to L1 English learners of L2 Arabic, and benefits shown by Showalter and Hayes-Harb (2013) with regard to L1 English learners of L2 Chinese. As Hayes-Harb and Barrios (2021) remark "we are aware of no studies to date that have investigated the communicative consequences of orthographic input; that is, the influence of orthographic input on the degree to which learners' speech is understood by others (i.e., intelligibility) or listeners' assessment of the ease or difficulty of understanding learners' speech (i.e., comprehensibility)" (p. 322). It is clear that in spite of the diverse knowledge based on isolated investigations of L1-L2 pairs, much more research is required on orthographic effects to enable further developments in optimal practices for language teaching and learning.

Global Englishes Language Teaching (GELT)

In the research literature, there are four relevant terms to conceptualise the ways that English is used transnationally with particular regard to listening: World Englishes (WEs; Kachru, 1989), English as an International Language (EIL; McKay, 2018), the newer umbrella term GEs (Galloway, 2013; Galloway & Rose, 2014; Rose & Galloway, 2019) and English as a Lingua Franca (ELF; Jenkins, 2000; Seidlhofer, 2011). Many of these distinctions have important epistemological and ontological differences, and yet there is considerable overlap. As already stated, GEs has set out to be a more inclusive term, and includes all the other paradigms. Rose and Galloway (2019) provide the best summary of the differences between the terms; essentially WEs has been used as a term to describe the side of the field which describes different existing varieties, EIL looks at the use of English for and in international interactions, and ELF builds upon this by taking into account accommodations and situational uses. All of the terms are used increasingly interchangeably as research knowledge coalesces. In this section, I

look at these terms under the umbrella of GEs and explain their importance for the context of phonology teaching in GELT.

McKay (2018) states that there are overlaps between EIL, WE and ELF. Both EIL and WE are concerned with content. According to McKay (2018), WE is concerned with cataloguing the different features of largely regional varieties of English throughout the world. It therefore contrasts with ELF, which is concerned with communicative interaction between interlocutors and how they accommodate each other. Rubdy and Saraceni (2006), however, highlight concerns that EIL and ELF could become a new de facto standardisation. In contrast, McKay (2018) states that "EIL scholars also recognize that what language is used and how it is used is a factor of both the purpose of the communication and the speaker's first language, culture, and level of expertise in English" (p. 11). As a result, it could be argued that EIL, while potentially useful as a tool of globalised neoliberalism and neocolonialism (Kubota, 2015), can also be considered from the perspectives of individuals as a means of liberating oneself from oppression. This is certainly the case with some L2+ users of English (Jones, 2024b).

As mentioned above, GEs is not actually a specific framework, but is a catch-all term which covers WEs, EIL and ELF, which had differing and even conflicting perspectives at times in the case of WEs and EIL. Kirkpatrick (2021) cautions that few empirical studies have examined GEs/ELF, and where they have, there has been not follow-up to ensure that positive findings are robust. Due to GE and GELT not actually being a framework Bayyurt and Selvi (2021) note that the local context needs to be taken into account, particularly regarding the varieties to be taught. It is unnecessary for speakers to be able to mimic speakers of other varieties, but phonology is useful for teaching listening comprehension (Jeong, 2021). Meer (2021) notes that in the German ELT context "most curricula do not systematically link audio-/audio-visual comprehension competencies to specific Englishes, but rather well-known standardized varieties of English and variation in English more generally" (p. 94) and that "varieties of English are almost exclusively approached on a very broad, abstract, and unspecific level that leaves considerable room for interpretation" (p. 99).

One of the key points that has underlined the studies in this dissertation is their situation in an EMI department in a Japanese university. Galloway and Rose (2018) note that such contexts are increasing, and therefore with them, so does the use of ELF. Ruegg (2024) notes that a considerable number of Japanese domestic students also enrol in EMI programmes, and therefore, if ELF is going to be successful,

accommodation is required not only on the part of the speaker but also that of the listener. While Kirkpatrick (2021) comments that the idea of 'native speakers' being the best model seems to be difficult to shake off, the context of having a multilingual student body whose primary shared language is English, the language of instruction, and with a multinational faculty should make the irrelevancy of native models obvious.

As ELT appears to be moving toward a greater use of GEs (Galloway & Rose, 2014; Rose & Galloway, 2020) and ELF (Jenkins, 2000), it is clear that learners will encounter a greater diversity of English varieties used in both recorded and live speech. While learners may claim difficulties with unfamiliar accents, the difficulties faced are no different than those faced when provincial 'native-speaker' accents, or even ungraded speech are used (Jones, 2024d). Authentic language use, whether standardized Englishes or GEs, are likely to be difficult for learners due to unfamiliarity and the prevalence of highly graded versions of standardized varieties in global ELT materials and standardized tests. Additionally, if we consider the use of the L1 in parsing the L2, due to activation of languages as triggered by phonology, we may also uncover different factors that may be useful in teaching English listening. This is particularly true for those teaching ELF (see Jenkins, 2000; Seidlhofer, 2011), where it is plausible that a listener may encounter an interlocutor with an unfamiliar L1 background which uses some allophones that, while comprehensible as L2, may initially activate L1 parsing. Obviously, it would be advantageous to teach strategies to avoid this stalled parsing occurring as much as possible, but it may actually happen on a subconscious level. Unfortunately, research in this specific context is lacking. Lévy and Grosjean (2008) and Gonzales et al. (2019) appear to be the only research into language activation in bilinguals/multilinguals.

The scenario above may arise for bilinguals and learners, according to the L2LP model (Escudero, 2005) due to a boundary shift where different phonetic cues cause phoneme mappings to overlap in cases where there are many-to-one mappings between L2 and L1 phonemes (Nagle & Baese Berk, 2021). It appears that using a wide variety of speakers for phonological exposure would provide greater latitude in the encoding of phonological prototypes (c.f. Kraljic & Samuel, 2007), thus enabling listeners to perceive phonological details more easily than from so-called 'native-speaker' input alone. As Tyler (2019) affirms "it is not necessarily a lack of native-speaker input that may reduce the likelihood of L2 category acquisition in the classroom, but input that fails to provide clear phonological differences between L2 categories" (p. 616). As such it is *not the case* that certain types of L2 speakers should be avoided because they may provide poor input. Instead, rich input needs to be

provided by teachers and sought out by learners, with clear evidence of differentiation in vowel quality as it pertains to meaning and parsing.

Potentially, phonology is harder to teach than other language systems, although arguably one of the reasons it is so difficult to teach is its absence from commercial textbooks. While dictionaries frequently contain IPA charts of the phonemes used for pronunciation transcriptions of citation forms for each entry, language teaching textbooks fail to include much except for some truly basic concepts, and with fairly poor guidance on teaching phonology for either listening or pronunciation.

As part of Jones (2017), commercial textbooks were analysed for the use of phonology for listening. While phonology was not entirely missing, it was misclassified as pronunciation (despite missing any instruction for articulation), and presentation of phonemes and phoneme clusters was not systematic. Teachers, particularly novice teachers and those with little in the way of pre-service or in-service training, may rely on published materials as a suggestion of good practice or even as the actual curriculum (Guerrataz & Johnston, 2013). As such, poorly developed materials may cause a cycle of teachers teaching phonology poorly, unless they become particularly engaged with the scholarship of teaching practices and learner outcomes. Furthermore, the focus of any phonology activities in textbooks tends to be production oriented, with any phonology for listening relegated to awareness raising without further application. Unfortunately, the language varieties prevalent in textbooks tend to be prestige varieties (i.e. standardized varieties used in the United States, Canada, the United Kingdom, Ireland, Australia and New Zealand) (Rubdy, 2015), and often highly-graded versions of those varieties. As such, familiarity with authentic prestige variety speech will be lacking, and GE familiarity is likely to be non-existent. This is reflected in participant comments in Jones (2024d) stating that both GEs and prestige varieties were difficult to comprehend, with explicit comments regarding GE unfamiliarity.

It is not the case that phonology materials for teaching and learning are completely absent. There are a number of teacher-oriented books on phonology, such as those by Roach (2009), Rogerson-Revell (2011), and Walker and Archer (2024). There are also teaching materials available, such as Hancock and Donna (2012). However, these materials are somewhat specialized, as they focus on one skill (pronunciation rather than listening), and are therefore likely to be less commonly used than four-skill general English textbooks. Those four-skill textbooks arguably

orient teachers toward testing listening comprehension rather than teaching how to listen (Field, 2008). Moreover, globally published textbooks are marketed to several different territories regardless of the L1s of the learners there, and thus do not take into account the differences between educational contexts. In spite of global publishing, there is scant use of GEs and the current situation is one in which GELT requires integrated tasks which contribute to noticing features of authentic language, which ELT coursebooks have neglected (Tsantila & Lopriore, 2024).

Clearly, the reality of English(es) use is a pluralistic, one which does not correspond to the existing, overly simplistic model of a small number of standardised models dictating how one ought to use the language. Because standardised models are often used as a set of prescriptivist rules in teaching materials (Kuo, 2006; also, Alptekin, 2007 for a counterargument; Prćić, 2012) than depictions of how some speakers use English, learnersmay develop mindsets that are maladaptive oriented away from language study. The first of these mindsets is that their English skills are deficient: that learners' English use is not, and likely cannot be sufficient, and that they ought to aim for a 'native speaker' model, which is unlikely to be attained (Munro, 2021), although potentially for a variety of reasons beyond effort and exposure alone; alternatively, they may come away thinking that language education is not suitable for them, because they see that the models used in the texts are not representative of the societies and the ways of being that they know (Gray, 2012). These mindsets run counter to the credo of most language teachers, who want their students to feel

the same passion for the English language(s) that led them to their teaching career, and yet the materials that are frequently in use in classrooms are likely to foster such maladaptivism.

In addition, by limiting materials to 'native speaker' models, the ELT industry/profession does a disservice to language learners because they are not provided sufficient exposure to a wide range of English varieties, which limits their ability to understand interlocutors who may not speak the same overly graded version of standardised versions of British or American Englishes. In fact, regarding the acquisition of English phonology, it would make more sense to provide access to a wide variety of spoken Englishes in order to broaden phonological templates (Vihman & Croft, 2007).

It is crucial that media is chosen that reflects the needs of learners, both short term and long term. Although studies claim that learners express preference for standardised varieties, Watanabe (2023) notes that "the preference for US English

owes a great deal to its continuing use as the practical model in ESL classrooms" (p. 37). While so-called 'native speakers' are over-represented in the recordings that are provided in published ELT materials (Kiczkowiak, 2021; Vettorel & Lopriore, 2013), it may be the case that such speakers are necessary models for speech in some cases. One particular example is study abroad, where learners may need to become familiar with particular regional varieties of English. Furthermore, the internet facilitates greater communication between people across nations and languages, and global society is increasingly mobile for a number of reasons. Due to the absence in ELT materials of this mobility among various socioeconomic groups (Gray, 2012), teachers and learners are likely to resort to finding their own materials from internet media that enables them to learn from a wider range of Englishes.

Multimodality and phonology

Speech is a multimodal process: not only do speakers produce sound sound by making articulatory gestures, but there is also the visual nature of the articulatory gestures themselves which are perceptible by listeners. With regard to vowels, this can be seen in the rounding of the mouth and the level of openness of the mouth. Additionally, contextual factors such as hand gestures and other ways of using the body can convey emphasis and pragmatic meaning. By perceiving these modalities, both auditory and visual, listeners can comprehend speech and speaker intention.

With regard to straightforward modality of speech, there is very little in the literature comparing audiovisual (AV) and audio-only (AO) modalities. Navarra and Soto-Faraco's (2007) results suggest "that the integration of visual-gestural plus auditory information can produce a specific improvement in phonological processing" (p. 9) seen in improved reaction times in identifying phonemes. Sueyoshi and Hardison (2005) also found that higher proficiency learners appeared to comprehension. Hardison (2017) used gated audiovisual stimuli in comparison with gated audio and found a greater effect in the perceptual acquisition of stop consonants for the audiovisual condition. Due to the nature of the articulatory gestures, stops may be easy to teach with this method, unlike nasals, liquids, and perhaps differences between some fricatives, due to the obscuring of the gestures, such as the tongue being alveolar but the mouth being closed for some realisations of English /n/. However, Inceoglu (2019) found that lipreading ability among participants correlated to greater perception of French nasal vowels. Because nasal vowels are produced with release through the nose, it may be assumed that they have low salience. However, given Inceoglu's findings, it may be assumed that what is assumed to have low

salience may actually be more salient than common-sense intuitions. Additionally, Cross (2011) found that visual information in audio-visual news-based materials facilitated comprehension but that learners did not pay attention to talking-head segments to gain phonological information to increase comprehension further. Similarly, Suvorov (2018) found that test takers often paid no attention to visuals in tests when provided. In sum, there are signs that the use of articulatory gestures for segmental phonology can be beneficial for acquisition, yet attention may need to be directed to those gestures.

A further factor that needs to be considered is that different levels of attendance to visual information may be language-dependent. For example, Yazawa et al. (2020) informed participants that "they would hear sounds from different languages... although they in fact heard exactly the same set of stimuli" (pp. 567-8) and this resulted in their attending to vowel length when they believed they were hearing Japanese, and attending to formant quality when they believed they were hearing English. Although this is a case of language informing attendance to different auditory qualities, it may be that visuals are attended to when attempting to perceive one language rather than another, or may even be L1 dependent. For example, Hisanaga et al. (2016) found that L1 English speakers attended to audiovisual speech more efficiently than audio only, which was the reverse for Japanese speakers.

One novel way of minimising orthographic interference is by use of visuals to represent phonemic information. Thomson (2012) worked on vowel acquisition with visual correspondences to flags, though one obvious criticism is that there is neither semantic nor gestural information given with the flags, therefore the visual information is merely a matter of convention for vowel identification and therefore does not provide visual information that ordinarily coincides with the auditory information. It may, however, prove useful in avoiding use of orthography without resorting to use of IPA transcription, which may be time consuming for learners.

Lambacher (1999) showed that software can be used to show waveforms as feedback for learners' consonant production, though this is not particularly aimed at building up perception of phonemes for any other reason than producing a benchmark for production. Using pitch contours, Hardison (2005) and with wave forms, Motohashi-Saigo and Hardison (2009) also reported positive outcomes with learning of prosody and geminates respectively. While these studies have focused upon pronunciation, one should remember that perception is a necessary preliminary step. Additionally, if waveforms or pitch contours are used to facilitate the ability to

discriminate phonemic and/or phonetic information, the extent to which learners can transfer such skills beyond the lessons, needs to be considered. Waveforms are not displayed in face-to-face speech, nor in telephone conversations, for example; therefore, the use of novel visual aids may not always be practical nor of obvious utility to learners.

Modern computer-generated images are increasingly realistic but, may result in 'uncanny valley' effects, where there is sufficient realism to represent reality, but not enough to be convince a viewer that what they see is anything but verisimilitude. As such, animations may not be as suitable to facilitate the acquisition of phonology as videographic sound recordings. Additionally, even videographic recordings have their weaknesses, such as the potential for jitter in online recordings to result in McGurk effects (Jones, 2021), which are misperceptions of speech due to mismatching visual speech gestures auditory speech (McGurk & MacDonald, 1976). Essentially, in a recording of audiovisual speech, the visual articulatory gestures of speech are mismatched to the auditory signal, which results in listeners perceiving entirely different consonant sounds. Green and Kuhl (1989) investigated VCV stimuli with audio-only and audiovisual stimuli perception and found that visual information affected participants labelling of consonants. However, McGurk effects have not been observed with regard to misperception of vowels or vowel quality.

In a study of monolingual Australian English users listening to Sindhi consonants in clear and citation speech conditions, Fenwick et al. (2017) found that actually audio-only speech was more consistently accurately categorised than audiovisual speech, which they attribute to interference from the visual modality. They theorise that "It is likely that we attune equally to the corresponding native articulatory gestures that we can see" (p. 122) when perceiving speech from additional languages. This is supported by Kolozsvári et al (2019), who found a mismatch negativity in audiovisual CV stimuli between Chinese and Finnish samples, for which all samples were familiar to Chinese and only half were familiar to Finnish speakers. A difference in effect size was also found for audio-only versus audiovisual stimuli, with larger audiovisual event-related potentials, suggesting that audiovisual speech is processed more easily than audio-only speech.

Some interesting work has been done on language teachers' stated beliefs and use of audiovisual input by Sydorenko et al. (2024). In a questionnaire study. language teachers were asked about their use of audiovisual technologies for learner input. They reported that 75% of their sample reported using video without L1 subtitles or

L2 captions “very often or sometimes” (p. 13) and a small number of the sample reported reasons for the use of video as being to provide a challenge to students, or for assessment. Although Sydorenko and colleagues’ study does not provide information regarding which language systems teachers and learners use audiovisual input for, it is clear that language teachers are using video widely, and as such it is necessary to establish potential effects on its use.

Multimodality and digital materials

The internet is an obvious place to search for teaching and learning resources, with very accessible and often free materials easy to find. This is even more useful for multimodal materials, particularly video, with user-generated content being arguably as popular as commercial streaming sites, although used for very different non-teaching purposes. Ravelli and Van Leeuwen (2018) comment on the proliferation of user-generated content that defined Web 2.0 as being considerably less subject to gatekeeping as old media, in “relaxing the rules that formerly determined who is a legitimate producer and the nature of the texts that can be experienced by others” (p. 289). In using web video, this may also result in the question of what counts as a learning text as opposed to a teaching text? This is important because the more motivated or cognizant of their agency the learner is, the more likely they are to go beyond the teacher’s intentions and explore set texts in depth and also find related texts to support their learning on their own terms.

Another important consideration in the use of technology in language teaching research is that while the same media may be sent out, there are differences in how it is accessed.

> If a visual text can be accessed via one medium only, it is, as a physical object, the ‘same’ for each viewer. However, they interpret it, all viewers will be confronted with the same degree of colour saturation, the same amount of detail in the representation, the same degree of modulation in the colours, and so on. Different platforms, however, affect these aspects of image quality, depending for instance on the size of the screen, the number of pixels and so on, so that different viewers may encounter different physical objects (to a degree this was already possible with the television screen, where users could choose their preferred level of brightness, contrast, and so on).

(Ravelli and Van Leeuwen, 2018, p. 291).

Thus, while controlling and standardising (digital) media stimuli is the norm in laboratory SLA studies, classroom studies with low or no funding use bring your own device (BYOD) not only as a necessity but also as an aspect of classroom teaching that ensures ecological validity of research findings.

Clark and Paivio's (1991) theory of dual coding has often been cited as a reason for providing input in both auditory and visual modalities, e.g. providing words and diagrams accompanied by voice. The crux of the theory is that information can be 'reinforced', that is the uptake of new information becomes more likely because the spoken auditory input is reinforced by visual images and/or text. It must be noted that Clark and Paivio's theory was based on L1 learning environments and therefore may not be directly relevant to L2 learning contexts. However, Cross (2009, 2011) has examined the use of news videos in teaching L2 English in Japan and found that visuals aided uptake of vocabulary given in the audio to an extent. In contrast, when investigating test taking, Suvurov (2018) found that test takers tend not to look at video images when listening but attend more to the audio. This lack of attention to visual information may be related to Suvurov's earlier work (2008), where L2 English listeners were found to be distracted by visual images provided at the same time as audio input. It can therefore be said that attention to visual information may aid acquisition, yet the task parameters can also determine whether or not learners exploit such helpful information or merely find it to be a distraction. This use of visual information does not refer to orthographic information which has particular challenges and which are explained below.

The use of orthographic input in language teaching is widespread and the de facto assumption is that orthographic input is a good thing, probably because of the social prestige accorded to literacy. However, it is not necessarily the case that orthography is advantageous, particularly for L2 English. Showalter and Hayes Harb (2013) found that novel orthographic input for Chinese facilitated greater perception of specific features of those languages, a range of studies (Grimaldi et al., 2014; Mathieu, 2016) have found that English orthography interferes with L2 phonology acquisition and therefore also L2 auditory perception. Of particular concern are the findings by Bassetti, Cerni and Masterson (2022), that teaching of phonology acquired from orthographic input is highly impervious to corrective teaching after the fact. While avoiding orthography altogether appears to be impractical due to the prevalence of English orthography in the global linguistic landscape (Backhaus, 2007;

Landry & Bourhis, 1997) it would appear to be more beneficial to avoid explicit use of English orthography for listening and phonology teaching until a fairly complete acquisition of English segments and phonotactics (which phonemes can collocate or not) has taken place. This is particularly the case with English in contexts where the L1 is transparent in terms of correspondence between graphemes (orthographic letters and sequences thereof) and phonemes.

Visual salience

The part that salience plays is important because it can modulate the effects of input frequency. By drawing attention to features that may be less frequent in spoken language by increasing salience of relatively unsalient items, there may be an increased probability of those items being acquired. However, methods of increasing salience can have unexpected effects. For example, by using orthography to increase the salience of spoken language, it is possible, particularly with English L2+ to create interferences in a learner's phonology that can prevent the acquisition of neighbouring phonemes in the vowel space (Sokolović-Perović, 2019; Bassetti, 2023). However, in other languages, orthography can play a facilitating role for phonology acquisition (e.g. Showalter & Hayes-Harb, 2013; Barrios & Hayes Harb, 2020).

Visual salience is a factor in successful listening comprehension. Sueyoshi and Hardison (2005) found that for advanced L2 listeners seeing a speaker's face while listening resulted in better comprehension than listening to audio alone, while intermediate L2 listeners had their best comprehension when seeing a speaker's face and hand gestures while speaking. In both conditions, listening comprehension was tested with multiple-choice questions. However, this study did not investigate phonological acquisition, even though it may be the case that visual salience facilitated accurate decoding of the speech signal. Hardison (1999) also played matching and mismatched audiovisual stimuli to participants of L1 Japanese, Korean, Malay, and Spanish to investigate perception of consonants. Matched stimuli were perceived better than mismatched stimuli by all participants, although this was modulated by L1. The reason posited by Hardison is that the different L1s have different acoustic qualities in normal articulation of the phonemes.

Visual salience is attested not only for perception of phonological items but also for acquisition of perceptual phonology. Hardison (2005) used gating of bisyllabic words and perceptual training in audiovisual and audio-only modes with L1 Japanese

and Korean participants. Gains in perception were higher for both audiovisual and audio-only groups after training, but gains were marginally larger for the audiovisual condition, particularly with /l/ and /ɹ/ onsets. The same findings were obtained in Hardison (2017), which found higher gains for perception of gated words in audiovisual modality than in audio-only modality, again with L1 Japanese and Korean participants. Hardison (2005) comments that there was an effect on /l/ and /ɹ/ learning particularly related to adjacent vowels, suggesting that vowels themselves are potentially visually salient.

In Jones (2024c) the findings regarding visual salience were complex. There is possibly an effect in acquisition, but one which could not be observed in the post-test. By the delayed post-test, however, participants had made gains, particularly with more peripheral vowels. This acquisition lag suggests that not only is noticing (Schmidt, 1990) important, but that it needs time to 'gestate', or for the stimuli being noticed to be internalised. Furthermore, my findings build upon Hardison's (2005, 2017) work, which showed that gated audiovisual input through video works to facilitate phonology acquisition. My findings show that simpler methods than gated video may also have a beneficial effect to phonology acquisition, although more research is needed with larger samples and in other contexts to verify that this is the case.

As such, it is clear that increasing visual salience can play a part in instruction of phonology, however it is not particularly easy to raise salience quickly or easily in a non- laboratory setting. It may be advisable for teachers and administrators to use materials that are developed specifically with raising visual salience of phonological items in mind. Such materials may be developed by commercial developers, although this does appear unlikely; as such, it may be beneficial for institutions to provide time for educators who are skilled in video editing to develop materials specifically for the needs of their learners. It would also be useful for the language teaching and learning community if such materials were made available either for free or to purchase.

Feedback

One of the most frustrating parts of being a listening researcher is that the processes involved in conducting research are inconvenient. Listening is a receptive skill, and as such it is difficult to observe what is happening when learners are listening. Reading can be done aloud, or tracked with a finger, speaking can be recorded and analysed,

and writing is there on paper or the screen to see. As such, research on feedback is much more focused on those other more easily observed skills.

How feedback appears to work in language learning is less by mere conditioning, although there are also proponents of this. For example, high variability phonetic training (HVPT) is effective in facilitating phonology acquisition (e.g. Cebrian, 2019; Mora et al., 2022; Mora & Mora-Plaza, 2023), despite its being highly behaviourist and its use of non-words rather than words in the target language. Another way in which feedback may be useful is through focussing attention and active cognition to errors in perception (and production, although this is outside the scope of the dissertation). HVPT is rapid by nature, whereas naturalistic feedback on listening or production may not be. As such, it would be highly beneficial for language teachers and language learners if the mechanism of how feedback works were more clearly understood.

Lyster and Ranta (1997) found that teachers in a French immersion programme in Canada in subject classes or language arts classes in grades four to six tended to use recasts as the most common form of corrective feedback, but found that this was ineffective in leading to negotiation of form, peer repair or self repair, in comparison to elicitation requests, metalinguistic feedback, clarification requests or repetition. However, it must be noted that this feedback was given in reaction to learner output of productive activities, rather than receptive activities such as listening, the focus of this dissertation.

Lee and Lyster (2016a) found that L1 Korean learners made better perceptual acquisition of the /i-ɪ/ contrast when their explicit instruction was corrective feedback. The corrective feedback provided was in-person, and provided immediately. Lee and Lyster (2016b) then investigated /i-ɪ/ and /ɛ-æ/ contrast learning with L1 Korean participants with four feedback groups: target form repeated (target), non-target form repeated (non-target), both target and non-target form repeated (combined), and indication of a wrong answer (wrong) and a control group receiving no corrective feedback. Greater gains were observed in the target and combined groups for both contrasts, and NRV effects were found with greater learning in the /i-ɪ/ contrast for both trained and untrained words than in the /ɛ-æ/ contrast.

In CALL systems, Lambacher (1999) has illustrated how software can be used by learners to compare their own output with the speech of a teacher, or from media.

This software uses spectrograms, which is similar to successfully work by Hardison (2005) using pitch contours for prosody and Motohashi-Saigo and Hardison (2009) using waveforms for geminates. However, this may not be practical for some teaching situations because there may be insufficient time to educate learners on how to use waveforms effectively due to curriculum demands.

According to Mackey (2012), learners in a text-chat synchronous computer-mediated communication environment often failed to notice recasts. This is unlikely in a listening feedback situation, but whether learners notice anything beyond whether an answer was correct or not, and if this contributes to their phonological acquisition is yet to be addressed.

Additionally, whether the timing of feedback can play a role in this would be a useful contribution to the field. As Schrooten (2006) notes, if resources with feedback are "well-designed, flexible and under user control" (p. 136) they can allow learners to gain feedback that is adaptive to their answers, and ultimately each learner's inner syllabus. However, the limitations of providing feedback according to learner input is not just a problem of correct or incorrect but a psychological information gap between programmer/developer and learner/user regarding why the answer is incorrect and how useful feedback can be given (Schrooten, 2006). If we look at this in a teacher-mediated CALL/Mobile Assisted Language Use (Jarvis, 2015) environment, interaction between teachers and learners can, but do not always, solve this problem. Teachers may be unable to understand the root of learner difficulties, for example. Similarly, in an autonomous learning environment, this problem can persist and may require more learner input.

Chapter 3: Methodology

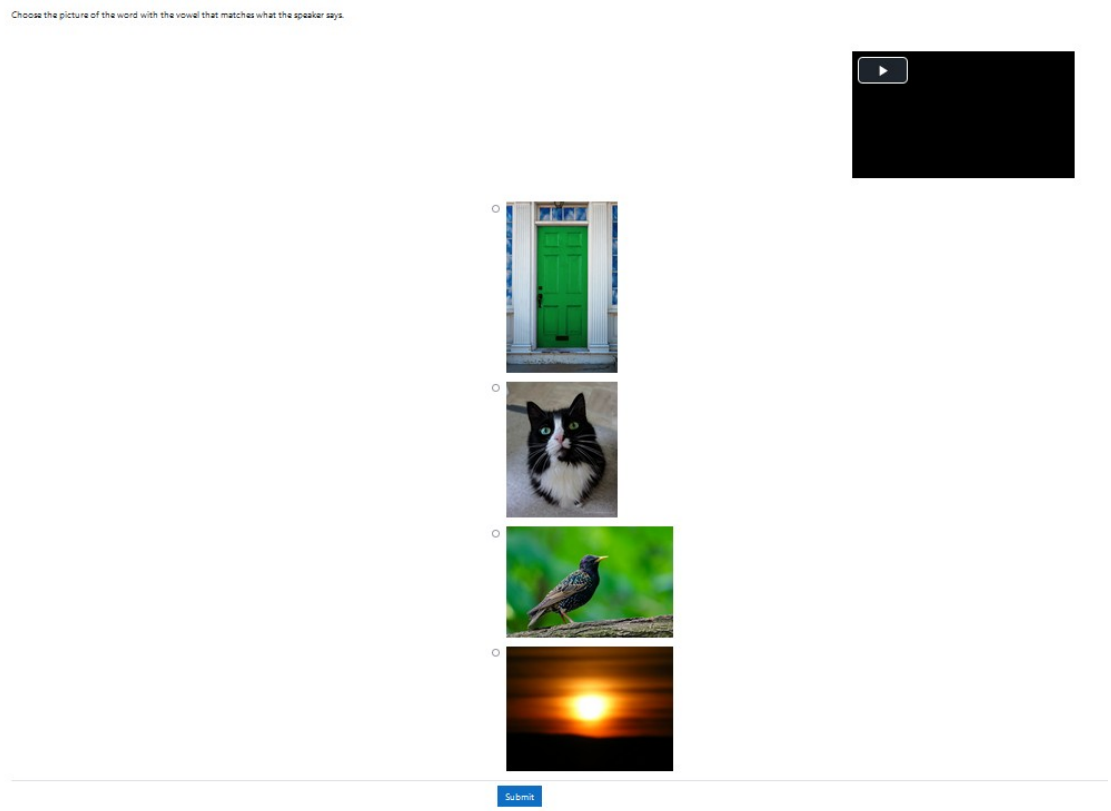
Data Collection

All three studies that make up this dissertation consisted of a pretest, which was followed by lessons conducted through the Moodle learning management system, and upon completion of all lessons, the following week a post test. One study, that on the effect of input modality upon vowel acquisition, also consisted of a delayed post test, taken four weeks after the first post test.

Each test consisted of 32 consonant-vowel-consonant English words, of which eight items each contained /æ/, /ʌ/, /ɔː/ and /ɜː/ vowels. These vowels were measured by using vowel identification tasks. Participants chose one example from a set of four pictorial representations of vowel types: pictures of cat, sun, bird, and door represented /æ/, /ʌ/, /ɜː/, and /ɔː/ respectively. These are all high-frequency words and it is guaranteed that all members of the target population know these words in orthographic form; only their vowel discrimination renders the words potentially difficult to parse in connected speech.

The reason for using vowel identification tasks was that the main alternatives are either too expensive or ineffective measures. For example, laboratory work done by Valerie Shafer and colleagues (e.g. Datta et al., 2018) has used mismatch negativity stimuli for event-related potentials, which requires participants to simply listen and an electrode array outputs electrical flow from across the brain, which results in a different potential as the participant perceives the different speech item. The equipment required is extremely expensive, prohibitively so for university departments that may not focus upon child language acquisition. Additionally, such equipment also requires training to use correctly. A more accessible method of testing perception is the AXB test, in which participants hear three stimuli in sequence and choose whether the middle example is the same as/more similar to the first or third stimulus. The AXB test is straightforward and has been used in many of the studies conducted by others that have informed the work in this dissertation (particularly Tyler, 2019). However, AXB tests were not used in my work due to frequent ceiling effects in pilot studies, which resulted in inconclusive results. Therefore, due to the unsuitability of both AXB and mismatch negativity tests, vowel identification tests without orthographic prompts were judged to be the best possible method to assess vowel discrimination.

Figure 2: Moodle screenshot for vowel identification in the test format.



Data analysis

The use of Bayesian data analysis, which should not to be confused with the use of Bayesian theory of cognition that has also been used elsewhere in this dissertation to discuss theory of phonology acquisition, is a break with the traditional use of frequentist statistics in language teaching research. This is as much of a philosophical choice as it is a pragmatic choice. The reasons for this are that the Central Limit Theorem does not apply, and additionally use of prior data informs the use of subsequent data (where available) to better predict the probability of an effect size being true. A further reason for the use of Bayesian methods is rigour: the use of null hypothesis statistical testing (NHST) is inappropriate in initial work that is intended to show proof of concept (see Trafimow, 2024 for further discussion of the misuse of NHST). The studies conducted that form the basis of the included articles were all novel work, conducted for the first time in classroom or virtual-classroom-based CALL environments. Stating that results are significant in a piece of initial unreplicated work is therefore unreasonable. Furthermore, the studies in the article also use small sample sizes, which means that the use of frequentist statistics is inappropriate due to

low statistical power and therefore inability to provide a suitable rationale for carrying out studies. Bayesian analyses are objectively more suited to small sample studies, which are the most prevalent in language learning studies because the Central Limit Theorem applies only to frequentist statistical methods but does not apply to Bayesian methods.

A further consideration of the choice of statistical methods is that of avoiding reliance upon p values. The p value is a measure of the probability of the resulting effect sizes being arrived at by chance. The larger the sample size is, the smaller the resultant p size. This can result in unscrupulous data collection practices, where researchers simply stop collecting data at the point where their p values reach significance (Stefan & Schönbrodt, 2023). Unfortunately, this type of frequentist reporting of statistics is often used in applied linguistics research. Even where researchers are well-intentioned, the use of p values, particularly for initial work, results in the continued overuse and overreliance upon p values. Beyond the weaknesses of frequentist analysis, the use of prior data to inform subsequent data Bayesian analyses is good practice, as use of data from other researchers' studies can assist in consolidating knowledge in our field. The use of Bayesian statistics is examined in the conference paper _Choosing Statistical Analyses for Small Samples_ (Jones, 2024a).

One important note regarding the methodology for comparing the GLMMs is that whereas Norouzian et al. (2019) use Alternative/Null = BF (p. 252), the method I have taken was actually the inverse, i.e. Null/Alternative, or comparing no interaction versus interaction. This means that where Norouzian et al. (2019) have Bayes Factors > 1 nominally supporting the alternative hypothesis, in my work they support the null, and where Norouzian et al. (2019) have Bayes Factors < 1 supporting the null, mine support the alternative hypothesis. This is a product of learning to carry out the GLMM model from a range of sources. The numbers are still robust; it is merely that the way of laying out the calculation and code is unorthodox. However, when a Bayes Factor is cited, it is discussed explicitly in regard to its support for the alternative or null hypothesis, or in the GLMMs interaction or no interaction.

Chapter 4: Overall findings

RQ1. Does a speaker's variety of English have an effect upon (a) vowel acquisition, and (b) ease of listening?

Phonology acquisition

The paper by Jones and Blume (2022), essentially says everything in its title: 'Accent difference makes no difference to phoneme acquisition'. The study is, to my knowledge, the only published work that takes a quantitative, quasi-experimental approach to GELT. There is an attempt at empiricism which is limited by the sample size, although this is mitigated by the use of Bayesian analyses as approached to the frequentism in most (but it must be noted, not all), ELT research.

The rationale behind the study and the article is that global mobility is increasing, facilitated by the spread of English as a frequently taught/used language worldwide, and yet the listening input that English language learners are exposed to in classroom materials is largely based upon Englishes from Kachru's (1988) inner circle. There are several studies on the conservative approach to the people featured in ELT textbooks, such as Kiczkowiak (2021), and Gray, (2012). Those studies have found that in global textbooks, the people are predominantly white people from America or the UK, or in Gray's study, rich people learning English for travel, leisure or other consumer-society related purposes. Those learner characters in the textbooks are usually featured in reading sections rather than listening sections. While there are attempts to include more speakers from outside the inner circle, these attempts are notable due to their rarity. Although some textbooks have featured voice actors attempting 'foreign accents' with limited success, textbooks such as *Navigate* (Roberts, et al., 2015) and Speak Out! (Eales & Oakes, 2023) have used authentic texts, or, in the case of online web materials aimed explicitly at English learners such as *ELLLO* (Elllo, n.d.) semi-scripted conversations to enable control over content while maintaining elements of authentic speech (Beuckens, private communication). Although such welcome changes are being made, they are slower to occur than they ought to be, partly due to inertia, and partly due to a lack of studies examining the differences between using different varieties of English. However, our hope was to explore how the different varieties of English affected learners' vowel acquisition within the conditions of how learners are likely to use English (with speakers of English with varied L1s), and the materials they may have used previously or supposed ideal models put forth (English from speakers of Inner Circle countries (Kachru, 1988).

The findings of Jones and Blume (2022), as mentioned in the title, show that there is no difference in vowel acquisition with regard to the varieties provided as input. The learners were assigned to two groups, either GEs or PVs, with assignments of approximately matched pairs across groups based upon pretest scores. Post-test gains by both groups were actually negative, therefore reflecting the fact that neither the PVs nor GEs had any notable difference in the phonological acquisition of the learners. One Bayes Factor in the GLMM was convincing regarding acquisition of the TRAP vowel (/æ/), at a value of 0.002, supporting an interaction between TRAP gains, condition, participant and interval standard deviation. However, the TRAP gain differences between groups in Jones and Blume (2022) was small, and this result may be due to noise in the data, because the standard deviation in the GEs group was much larger than that of the PVs group. Admittedly, some factors in the study may have had a much more significant role in gains.

Firstly, the intervals between lessons and tests for learners show that most of the learners had 'crammed' their lessons shortly before the deadline required for assigning academic credit to the work. Spaced learning has documented benefits for L2 learning (Kim & Webb, 2022), and therefore can safely be assumed to be more beneficial to language acquisition than cramming. Secondly, the learners may also have been distracted by notifications, etc., on their devices, and therefore lessons may not have been as subject to high-quality attention as assumed in the planning stage of the study. However, the findings still stand, on a preliminary basis. Real learners, with real learner behaviour, participated in the study, and therefore whether PVs or GEs result in higher vowel acquisition scores is shown to be conclusive in that the factor of input variety are irrelevant compared to other factors involved in teaching and learning. As such, Jones and Blume (2022) recommend that GEs be used in teaching and learning materials in order to rectify linguistic imperialism, while also bearing in mind learner preferences and target communities with whom they are likely to interact.

In order to facilitate the acquisition of English phonology among their learners, it would be useful for teachers to consider the L1 phonology and any additional languages their learners know (L1+) and find the overlaps with English. Additionally, investigating the level of incongruity between the L1+ and English orthographies would be useful in order to minimize possible orthographic interference and to maximize its efficiency. While seemingly counter-intuitive, minimizing mimicry with novel phonemes may be more beneficial in the early stages of language learning. Providing visual aid through video also appears to be beneficial, although learners' attention must be directed toward the speakers' mouths and articulatory gestures.

As mentioned above, the research is quite disparate in places, with several studies across L1-L2 pairs, with little acknowledgement of the multilingual reality of much of the world. More mixed-methods use (Cresswell & Plano Clark, 2018) in order to triangulate quantitative data with learners' lived realities, can provide answers not

only to the how and what of phonology acquisition, but also to the why. Additionally, more research on phonology pedagogy in classroom and/or institutional contexts can provide a more ecologically sound research base to impact practice and provide affordances for replication by teacher-researchers.

Ease of listening

Qualitative data was collected from learners in the same study that forms the basis for Jones and Blume (2022). This qualitative data provides the basis for Jones (2024d). Participants were asked to provide feedback regarding what they found easy or difficult to understand from the recordings they listened to.

Essentially, the findings state that GEs are no more difficult to understand than prestige varieties of English. Pronunciation and unfamiliarity with a variety can cause difficulties for learners, but such unfamiliarity and pronunciation factors are not exclusive to GEs, and in fact, the speaker most widely reported as difficult to understand by participants was a white American male TED speaker, not the type of person one would stereotypically assume to be difficult to comprehend. Furthermore, GEs speakers could be motivating to learners, and to encourage more engagement with the content of the recordings. This finding also appears to be similar to Chen's (2022) study in the Taiwanese higher education sector.

RQ2. Does input modality affect vowel acquisition?

Jones (2024c) found that learners acquired vowels more readily in the AV modality than audio only, albeit with self-selected learners who took a delayed post-test. These results are very weak evidence and the Bayes Factors are all weak (< 10) yet in favour of no interaction between conditions and acquisition. It may be a case that a lag effect is at play with visual salience, because the learners' gains only occurred markedly at the delayed post-test stage. However, the difference in gains between groups suggests that the effect is real. More research is necessary to investigate whether other factors, such as cramming, cause noise in the data.

RQ3. Does feedback timing affect vowel acquisition?

In Jones (2024e) I presented findings from a classroom setting where learners were provided with immediate or delayed feedback on phonology and listening exercises. Despite attendance problems caused by an outbreak of COVID-19 among the students at the institution where data collection took place, sufficient data was captured in order for tentative conclusions to be drawn based upon exploratory data analysis. The

conclusion made was that feedback timing appeared to have an effect upon vowel acquisition, with delayed feedback being more beneficial than immediate feedback. While the GLMM result for gains of all target vowels without relation to lesson performance was robust in favour of no interaction with gains, the models indicate that for each vowel, there is an interaction between gains, condition, participant and interval standard deviation resulting in Bayes Factors in the range of at the strongest evidence for STRUT at 0.0613003 to the weakest for TRAP at 0.1751258, which Norouzian et al. (2019) would class as substantial evidence.

Potentially the reasons for the difference in effect is that immediate feedback is subject to less attention due to task switching effects; that is, learners answer questions, then focus attention briefly upon feedback and then switch attention again to the next question. This point of view contrasts with that that of Kim and Webb (2022) who posit that simply more laborious processing leads to acquisition. However, in delayed feedback situations, attention may be paid to all sub-tasks, without the need to switch attention until the end of the task, where feedback for each question was made available. It may be the case that delayed feedback is beneficial only due to there being less need to regulate attention. Furthermore, feedback is not equally effective with all of the vowels taught, with more peripheral items subject to higher gains due to the NRV (Polka & Bohn, 2011). As with my investigation into visual salience, more research is required.

Additional findings

Cramming

An additional finding across the studies is that some learners from each sample had a tendency to cram their study rather than spacing it as intended. For some of the learners in the third study (on the effect of immediate versus delayed feedback), this cramming was often due to making up for absences. In the other studies related to GEs and modality making up for missed lessons was not the reason, but mainly that participants had not logged on to the Moodle website because they had procrastinated. In Jones and Blume (2022), there was a very large Bayes Factor (118.91) regarding the standard deviation of interval between lessons and the gains between pretest and post-test.

Cramming may be particularly suboptimal for language learning, based on the existing research on performance in L1 verbal recall tasks (see Cepeda et al., 2006 for a review) to frequent encounters with language items in different episodes of attending to form being particularly useful for acquisition that is robust over time. On the one hand, participant cramming in the studies led to noisy data. However, on the other hand, the fact that cramming was observed is not necessarily a negative for the studies; cramming is frequently observed student behaviour, and therefore it provides another factor contributing to the ecological validity of the studies. Additionally, it was

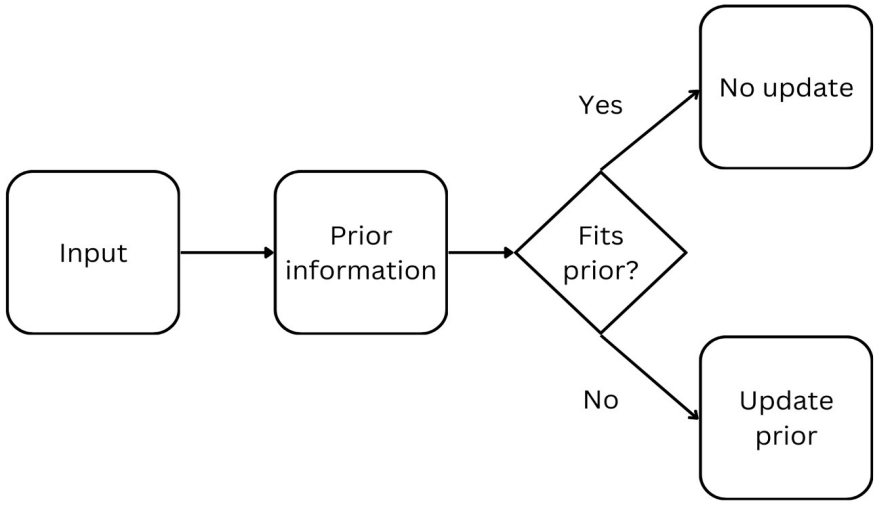
possible to control for cramming in data analysis by using the standard deviation of lesson intervals as a variable in the generalized linear mixed models.

Chapter 5: Theories developed

A Bayesian view of phonology acquisition

The way the brain handles language-related episodes shows evidence that it is broadly Bayesian: It makes predictions based on plausibility when it has prior information, otherwise it avoids decisions or makes random decisions (i.e. uses a flat prior). Upon receiving evidence, the posterior becomes the prior and the cycle continues. (c.f. Faerch & Kasper, 1980; Norris et al 2003, cited in Cutler 2012, pp. 396-8). According to Jakobson and Halle (1956) the probability of a phonetic feature occurring depends upon the feature itself and the type of text. If a stream of speech is heard, it is possible "to predict with greater or lesser accuracy the succeeding features, to reconstruct the preceding ones, and finally to infer from some features in a bundle the other concurrent features" (p. 5). This opinion is therefore commensurate with a Bayesian view of language learning, particularly input. However, for L2+ learners, their models of occurrence probability (i.e. their priors) are based upon L1 knowledge, or else upon incomplete L2+ knowledge, and therefore potentially lead to incorrect predictions and reconstructions.

Figure 3: A flow diagram illustrating the Bayesian cognition of phonology acquisition.



Ellis (2002, 2006) also states a belief in a Bayesian model of language learning cognition. As a model of phonology acquisition, the Bayes model is convincing because, as exposure to the L2+ occurs, phonological information is categorised based upon prior information: either L1 information or other previous L2+ information, or as Faerch and Kasper (1980) conceptualise it, from "linguistic experience..., communicative experience... [and] language learning experience" (p. 70). The more

information that is gained, the more informative the prior model of L2+ phonology, which leads to more accurate judgement of perception.

The phonology model, whether it is based upon the PAM (Best, 1995), the SLM (Flege, 1995), L2LP (Escudero, 2005) or ASP (Strange, 2011), is compatible with such a Bayesian model because all are based upon both L1 and L2 information. While the PAM (including PAM-L2, Tyler & Best, 1997) states that there is only one phonological template, it is still subject to both L1 and L2+ categorisation. The SLM (and revised model) (Flege & Bohn, 2021), states that there are phonetic rather than phonological categories but they are stored in models for each language. In this case, the L1 model is the prior and the L2 is the model of new information, and will be perceived according to the prior model and cumulatively contribute toward it. The L2LP (Escudero, 2005) is also compatible with a Bayesian model. Again, as with the PAM and SLM, the L2LP models language learning where the learner has already acquired L1, which again, forms the default prior model. However, L2LP states that the L2 is built on a copy of the L1 and developed accordingly. In this case, the prior model stays as is and the learner may access one of two Bayesian models — a default L1 prior, or a developing L2 prior, with any further input being added to the latter.

An interesting point raised by Baratta (2019) is the perceived need for a standard variety of English to use as a model in the classroom, with other varieties also provided. While this seems to provide a sort of linguistic pluralism, it in fact is simply another form of prescriptivist and a way to accord inner-circle Englishes more prestige. One way to provide an alternative to the hegemony of convenience is to consider the model of how language is acquired in a connectivist, usage-based model, language is acquired by exposure to linguistic evidence and frequency of occurrence, causing psychological cueing effects (Ellis, 1994). The brain makes sense of this through Bayesian means of considering likelihood of occurrence based upon evidence so far, and all future evidence is assessed based on what has also come before; that is, the model is in constant flux, constantly being updated according to new stimuli/evidence received. While this may become more complex as different domains (places and contexts) and tenors (register and interlocutor/audience factors) are considered, exposure to them in media and simulations may be sufficient for L2+ acquisition as may be the case with L1 acquisition.

Accelerating phonological acquisition through visual salience may require a communicative context, either interacting with an interlocutor or comprehending audiovisual media in the target language. This provides affordances for learners to test linguistic hypotheses, particularly regarding phonemic categorisation, through ongoing efforts in parsing and comprehension. This testing of hypotheses is likely to occur through visible gestures of articulation being perceived, which in turn cue phonetic phenomena to be parsed as phonemes. This would be based on L1 categories initially (pace Best, 1995). Feedback may be obtained through a sympathetic interlocutor providing assistance in comprehension, similar to Long's (1983)

Interaction Hypothesis; alternatively, if media are comprehended satisfactorily from the learner's point of view, this is equivalent to positive feedback, and unsuccessful comprehension is equivalent to negative feedback. However, with equivalent negative feedback, repeated attempts at comprehension are probably required in order to categorise a phone as a phoneme correctly/accurately.

The origami theory

As an analogy to consider the mechanism of phonology acquisition through perception, I would like to posit a theoretical model, which I have called the origami theory. Consider the auditory motor cortex as analogous to a piece of origami paper. The more that piece of origami paper is folded in particular locations, the more likely that folds will start to develop in those places because the paper becomes stressed with use. Without use in other areas, the paper does not fold so easily there, and remains pristine. With the audio motor cortex, this use principle applies to both perception and production, albeit with some physical limits such as auditory capacity and articulatory limitations such as tongue length and muscular capacity, for example.

Another analogy could be walking on a lawn, where frequent footsteps over certain routes result in erosion of grass, creating a hardened earth path. This path becomes more and more used over time due to its high visibility and also its likely convenience in taking a walker from point A to B. This is similar to the categorisation of phonemes, where more frequently used phonemes make a greater impression, whereas less frequent and/or less salient phonemes make a lesser impression and are less easily categorised, resulting in lower likelihood of perception while listening.

If there are places in the auditory motor cortex that are subject to wide variation in L*x* but not L*y*, then there is likely to be a wider acceptability of phonemic categorisation for L*y* until L*y* is acquired, with its own set of parameters, which would occur only after a certain number of encounters (due to need to retrieve and assign those parameters). Production is similarly constrained, due to L*x* parameters initially, and likely to result in a typical L*x* production, again until L*y* parameters have been set, although these are also constrained by articulatory limitations as mentioned above. The way that these limitations may manifest is given in the example below. As Meierkord (2012) points out, in situations when interlocutors use ELF and accommodate to one another:

> When a particular pool of features is available, it is the individual interlocutor who is making choices from among the competing features. Potentially, s/he may choose from all the features available in the pool. For one of the features to eventually become selected, the individual needs, first of all, to notice the feature, which may be a conscious or unconscious process. In case an individual does not notice individual features, these are not available to her/him for choice. (p. 64)

If there is a parallel process to this speaking accommodation in GELT listening, then this means there is no 'choice' before language acquisition, in other words, and listeners can choose only the phonemes that they perceive. How listeners 'change' the decoding features to in order to parse speech correctly when a speaker has an unfamiliar accent or dialect is, as yet, unclear, but likely relies on particularly salient features of input, for example a word that may contrast with an existing known word and which requires parsing with little contextual information.

Phonological memory and language aptitude

The way that phonology may be particularly important for language acquisition, particularly L2+ acquisition, is that development of phonology, particularly phonological memory and the ability to access phonetic contrasts, is likely to play a part in other aspects of acquisition such as vocabulary recognition and acquisition. Sparks and Ganschow (1991) explicitly make the case for lack of phonological awareness (across all languages) contributing to lower proficiency in learning an L2, including both reading and listening comprehension. It may be the case that learners who have better phonetic discrimination aptitude make better phoneme categorisations in L2+. These categorisations in turn lead to better word recognition, and ultimately better comprehension (Lange & Matthews, 2020; Cheng et al., 2022).

L2 working memory capacity has been found to correlate with L2 TOEIC listening scores, but only among lower proficiency learners, while higher proficiency learners' scores correlated with vocabulary (Satori, 2021). However, the literature indicates that ability to parse more accurately does not necessarily increase with only exposure to the L2+ (Carroll, 2017; Kenanidis et al., 2023). This suggests that explicit learning is a necessary condition, or that attending to salient aspects of the L2+, at least for adult L2+ learners.

Cueing also has an effect on phonological acquisition. When phonemes are learnt, they are learnt in context, with other neighbouring phonemes creating phonotactic cues based on exposure frequency in relation to their phonotactic collocation (i.e. the frequency of their occurrence adjacent to one another). As such, when one item cues another, it is an automatic process, and therefore does not require much if any working memory capacity for the cueing to occur. As such, learners who have received a large amount of input, with affordances for attending to the input (such as raised visual salience and delayed feedback, which fosters greater attention), are likely to develop better, more automaticised L2+ parsing abilities, based upon cueing and also phoneme discrimination abilities, which are based upon phoneme categorisations.

Chapter 6: Discussion

Contributions to research

The studies conducted as part of this doctoral dissertation make important contributions to the research community.

1. First of all, what works and what does not work with regard to instructed phonology acquisition on the whole, and perceptual acquisition in particular, is still somewhat unknown. These studies shine a light on an otherwise dark place. One particular area which the studies illuminate is that the findings support Polka and Bohn's (2012) Natural Referent Vowel framework regarding the learnability of more peripheral (extremes of 'open' or 'closed' and tongue height) vowels in the vowel space.

2. Second, because the studies use small samples and use analyses that attempt to mitigate this perennial problem in our field, they provide guidance to other researchers who may have similar constraints. Additionally, the data analysis files provide a potential template for researchers new to using R to be able to clean and analyse data downloaded from Moodle installations.

3. Third, in making data cleaning, data analysis scripts and data available, the use of Open Science practices by a relatively Early Career Researcher adds more pressure to the applied linguistics research community to adopt these practices as a default. If a doctoral student can adopt OS practices, there is no reason why an established researcher cannot adopt such practices. Furthermore, as more and more studies are published with data and analysis code, the research becomes 'future proofed' against being categorised as being in the methodological Dark Ages.

4. Fourth, because the studies were conducted with intact classes they are ecologically valid. Laboratory-based studies are certainly valid, of course, but upon transfer to the classroom there are often problems or additional considerations to take into account. By using the classroom as a research site, the considerations that are necessary are essentially the same as those of teachers who need to facilitate the phonology acquisition of their own students. The findings of the studies also come some way to meeting the suggestion of Van Engen et al. (2022), who state that "A more productive line of research requires that we focus on audiovisual speech comprehension in situations that are closer to real-life communication" (p. 3223). Arguably the classroom is an authentic place, and learners need to develop listening skills in order to participate in lectures and other academic activities, yet the findings may go even beyond the classroom context.

What factors affect perceptual vowel acquisition?

In Jones and Blume (2022) we found that neither prestige varieties nor Global Englishes have an advantage for the development of vowel acquisition in L2+. Additionally, as found with the same sample in Jones (2024d), neither the GEs group nor the PVs group had a particular difference regarding statements of ease or difficulty. These findings clearly spell out a case for the use of Global Englishes over prestige varieties in order to right wrongs over the neglect of Global Englishes due to the native speakerist hegemony based entirely on convenience that has pervaded English language teaching.

Regarding visual salience as a means to facilitate phonology acquisition, Jones (2024c) provides contextual evidence for audiovisual information facilitating phonology acquisition. However, this is not a simple linear relationship, and potentially involves a time lag. The reason for the lag in acquisition is possibly due to a need for the stimuli information to be connected to existing information and applied to further input (either deliberate or incidental) to update internal hypotheses, which remains compliant with the Bayesian model of phonological processing posited earlier.

One further finding regarding visual salience is that there appears to be a time lag apparent in vowel learning gains differences between audiovisual listening and audio-only listening, as found in Jones (2024c). The gains for both audiovisual and audio-only groups were fairly similar at post-test. Only in the delayed post-test was there a difference in gains. These gains appear to be modulated by the NRV (Polka & Bohn, 2011). However, whether the time lag is an artifact in the research due to participant attrition between post-test and delayed post-test or is a real effect requires further research.

With regard to feedback, delayed feedback appears to be more effective in facilitating vowel acquisition. It may be the case that greater attention can be paid to delayed feedback, and therefore the information in the feedback is more likely to be integrated, whereas immediate feedback may only be attended to superficially while learners are mentally preparing to work on subsequent sub-tasks, and as such are at least partially distracted.

The findings regarding the benefit of delayed feedback over immediate feedback are in accord with the Canals et al. (2020), but are different to the SLA work on vocabulary learning. It therefore appears that, while vocabulary learning and phonology acquisition can be mutually beneficial (see, for example, Bundgaard-Nielsen et al., 2012), they are not part of the same system, and therefore L2+ phonology development does not necessarily occur as a by-product of orthographic and/or semantic recall of lexis. On the other hand, the greater effect on vowel learning for delayed feedback may simply be due two factors: if the feedback can be repeatedly

played, then this results in repeated attention to the feedback; additionally, if the type of task that a learner is engaged in rapidly switches (i.e. identifying vowel tokens and attending to feedback), this may result in less attention to the details of the feedback.

Work in the dissertation also adds support for the NRV (Polka & Bohn, 2011) in an instructed SLA context. Jones (2024c, 2024e) and Jones and Blume (2022) found that the /æ/ vowel was more readily acquired and that /ɜː/ appeared to be less readily acquired. /æ/ is the most peripheral vowel of the target vowels instructed in the studies (/æ/, /ʌ/, /ɜː/, and /ɔː/), and /ɜː/ the most central. The discriminability of /æ/ is unsurprising, because it is so close in the vowel space to Japanese /a/, and therefore is categorisable as the same, in much the same way that L1 British English users categorise American English /a/ as British /æ/ because there is no functional difference. As Polka and Bohn (2011) posit, the peripheral vowels are essentially universals and thus function as references for the other items in the vowel space.

The NRV may also relate to the effects of visual salience for perceptual vowel acquisition, because the peripheral vowels are at the extremes of open/closed and back/front and potentially also spread/round (although the studies in this dissertation use only one rounded vowel, /ɜː/). These extremes of articulation are easier to perceive visually, and therefore may have a greater impact on the learnability of certain vowels. It is also particularly useful to note that in Jones (2024c), the audio group learned /ɜː/ more easily than the audiovisual group. This difference in learning may be due to lower interference between visual and acoustic information; that is, for L1 Japanese learners of L2+ English, /ɜː/ may be perceived more easily through a focus on formant information alone, whereas, for example /æ/ is more easily perceived through a mixture of acoustic and visual information.

Small sample methodology

The use of small samples has been identified as a persistent problem in in applied linguistics research (Lindstromberg, 2023). Recruitment of participants is therefore difficult and requires goodwill on the part of participants, frequently the students of a teacher-researcher as in all of the studies included in this dissertation. Compounding the recruitment of participants is the fact that not everybody who agrees to participate in a study sees it through to the end, often in spite their intentions, therefore taking participant attrition into account is not only necessary but should be assumed to occur. If a minimum sample size for statistical power as per frequentist statistical norms cannot be reached, a change of methodology or analysis is necessary. Many methods are available for quantitative research with small samples (see van de Schoot & Miočević 2020 for examples), so it is not necessarily the case that quantitative researchers need to pivot to qualitative methods.

However, even beyond this use of small samples, an idea was put forward in Lowe and Jones (2024) about considering research as an ecosystem, suggesting that

researchers consider a plurality of approaches. Certainly, one of the weaknesses in our field is of 'novel' research using frequentist null hypothesis statistical testing (see also Trafimow, 2024), when actually, if research is truly novel, it would merit an analytic approach which is exploratory rather than confirmatory.

Open Science practices

Importance of code

If interested parties do not have access to the code, they cannot see how the statistical analyses were actually carried out and therefore open data is nigh on useless. There are a number of software packages available that can be used to perform statistical analyses, the most well-known of these likely being SPSS (IBM, n.d.). However, SPSS files lack interoperability with other software. On the other hand, SPSS is deemed to be more user-friendly than some of the open-source software that is available.

In the Open Science community, R, Julia and Python tend to be frequently used, although these are actually programming languages. However, a large number of code libraries and packages exist to enable researchers to create their own statistical analysis depending on the form that their data takes.

For these studies, I used R (R Core Team, 2022) within R Studio (Posit Team, 2024) as my software environment because it is the most prevalent in phonology research and also because it is interoperable with Jamovi (Jamovi team, n.d.), an open source software package that I used in my somewhat recent Master's degree programme prior to doctoral studies. In my adoption of R, I have made my data analysis more open to the research community and also made it possible to future proof the analyses against obsolescence by using Simonsohn and Gruson's (2021) groundhog package, which uses R versions and library packages that were current at a defined date and which can be retrieved by any third parties.

Contribution to pedagogy

The contributions to language teaching pedagogy within this doctoral dissertation, and especially English language teaching pedagogy, are a negation of the 'native speaker' as model, which was found in Jones and Blume (2022) to not matter for the purposes of phonology acquisition, and in Jones (2024d) to not matter for the purposes of motivation nor for ease of listening. It is important to acknowledge that not all speakers provide useful input; low-proficiency speakers may be simply incomprehensible and demotivating to listeners. However, speakers who can communicate clearly provide a reason to listen, and perhaps also a model for learners to emulate, whereas speakers from Inner Circle countries may be models which are intended but ultimately dismissed and any motivation learners have to listen is more likely to be related to content than English variety spoken.

A further contribution to pedagogy is finding in Jones (2024e) that delayed verbal feedback is more likely to contribute to phonology acquisition than immediate verbal feedback. This finding may be counterintuitive for many teachers, who may be tempted to correct misidentified phonemes immediately. Certainly, for many web apps and software applications, particularly Duolingo (n.d.) as perhaps one of the most widely-used language learning applications, immediate feedback, albeit multi-modal, is normal. Whether multi-modal immediate feedback is more likely to result in greater learning than delayed verbal feedback remains to be seen, as does whether verbal feedback is also better than immediate feedback for other language systems such as syntax, semantics or pragmatics.

Beyond the above, two tentative findings from Jones (2024c) may also inform language teaching pedagogy. The first is that audiovisual representations of speech are more likely to lead to perceptual vowel acquisition. The second finding is closely linked, being that acquisition may not be immediate and may require time to be applied to further input. As such, it is hoped that teachers and materials developers make more of the affordances that video can provide for phonology input. However, it is also hoped that teachers and learners do not expect instantaneous phonology gains but realise that learning a language is an endeavour that requires a considerable amount of time, and in fact continues across one's lifetime.

Perhaps it is symptomatic of our times that expectations concerning time of many everyday activities are increasingly oriented toward instant fulfillment or instant satisfaction. However, in attempting instant learning in language pedagogy what may actually result is counterproductive. With a more patient outlook both at the lesson scale and the course scale, where cognitive processes and subconscious processes are given adequate time, perhaps language learning can occur without a somewhat neurotic preoccupation with learning gains over arbitrary time frames which do not map to cognitive realities.

Limitations and further research

One consideration that must be considered is that different language pairings may give different results. This can be due to various factor such as amount of phonological similarity/overlap in phonological inventories, as well as salience due to differences in use of suprasegmental phonology. Additionally, in locations where the L1 is largely confined to the population under consideration (i.e., in Japan, where Japanese is not widely spoken as L1 outside the country), cultural considerations may actually confound the language factors. One example of this may be the orthographic effect between Japanese L1 and English L2+, which results from English's larger phoneme inventory, and also the use of English orthography in language teaching prior to learners' sufficient acquisition of English phonology for listening. While many different settings may use orthography early in learning, some of these settings, such as in many European settings, share an orthography and may also have an equally

large, or larger vowel inventory. Therefore, before applying findings of this research or other related studies, the original studies' contexts as well as the context to which the research findings are applied must be considered and evaluated for compatibility.

The articles discussed in this dissertation were written up as studies of learners with relatively advanced (orthographic) morphosyntax but with low-intermediate morpho-phonology, particularly regarding vowels. More research is required into whether/how such learners can unlearn deviant phonology and acquire a phonology that enables greater understanding of a variety of Global and/or Regional Englishes. Note also that what I refer to here is not deviant pronunciation but an internalised perceptual phonology that lacks sufficient vowels categories in its inventory to accurately perceive target language speech from a range of speakers across different varieties including both prestige varieties and Global and Regional Englishes.

A further consideration is also the choice of vowels in the study. As already mentioned, the articles support the Natural Referent Vowel framework (Polka & Bohn, 2011), which states that more peripheral vowels are more easily learnt than more central vowels. The articles I discussed use the vowels /æ/, /ʌ/, /ɜː/, and /ɔː/. However, different vowels with different peripherality and also different distances from L1 vowels may give different results. More research in this area is necessary, but beyond the scope of this dissertation; the target vowels were selected for the difficulties they cause learners in their listening.

The articles discussed here are also investigating, in general, L2 acquisition. However, different outcomes may be possible if the studies were undertaken to investigate L3+ acquisition. L1 categorisation informs L3+ phonology acquisition, and furthermore L3 affects L1 and L2 phonology (Gut et al., 2023), and therefore adds further factors into an already complex system.

Cramming effects are also present to an extent across the different studies which are discussed in the articles. The participants did what many students in many contexts do: they procrastinated and completed work at the last possible minute. Such procrastination is highly probable to have resulted in lower gains than would otherwise be possible. These cramming effects were controlled for in the GLMMs as much as possible. However, with higher participant numbers (or lower rates of attrition), it may be possible to observe clearer differences in the gains (or otherwise) between conditions.

An additional limitation of the articles in this dissertation is the use of devices to view and listen to the materials. The learners participating in the studies discussed used a range of different devices to take the tests and complete the lessons, from smartphones to tablet computers to laptop computers and even gaming laptops. This difference in devices may have result in differences in attention paid to any visual

stimuli on screen. However, post-COVID, a BYOD culture has accelerated device use in Japanese higher education, and learners will use devices with which they are comfortable. A related point, and more important, is also the listening equipment with which participants accessed the materials. Some students used noise-cancelling headphones, other simple closed-back headphones, others used inner ear headphones and others who had forgotten to bring or charge their earphones or headphones used the built-in speakers on their devices. This is far from ideal but it is highly representative of how students attempt to access multimedia language learning materials at the present time. While the findings are transferrable between BYOD contexts, it is certainly possible that other technology-enhanced language learning settings would see less transferability (Pawlak & Kruk, 2023). Furthermore, in Jones (2024c), when comparing audiovisual and audio-only modalities, it is obvious that in an EMI programme, participants would be exposed to natural audiovisual English stimuli outside the class environment used for the study. This is to be expected, and as such, the gains for AV group are actually related to the salience raised for the input in that particular study and are therefore robust.

In summary, while the findings of the articles are valid and transferrable, they are not universally transferrable. Context is important to consider and therefore, as the findings have importance, they may be affected by language pairings, instructional context, learner proficiency, device use and also individual differences. Again, emphatically, while the findings are quite robust, the context to which they are transferred to introduce further factors into the phonology acquisition system.

Plans for further research

As with all research projects, the adage that more research is required applies. The directions that this further research may take are outlined below.

First of all, the use of GEs and prestige varieties making no difference should be verified with other groups of learners, perhaps over a longer period of time. If this could occur, as part of a replication or a partial replication, it would be useful for the field not only conceptually but also politically. Regardless of being a language learner or not, language users need to be able to understand a wide range of speakers. People want to comprehend interlocutors and they want to be able to comprehend media that they view or listen to. Phonology acquisition is the first step to comprehension. As such I plan to undertake further research and encourage collaboration between teachers/researchers in order to foster crossover between teachers and/or researchers (and teacher-researchers) working with different investigation and analysis methodologies, both quantitative and qualitative.

Second, the difficulties faced by L2+ listeners with Global Englishes and prestige varieties should also be studied more empirically. It would be beneficial to take preliminary data as in Jones (2024d), and triangulate this with stimulated recall,

which could a provide richer data regarding particular listening difficulties and the contexts and conditions under which those difficulties arise.

Third, regarding visual salience, a longer study with a larger sample is required, particularly to observe whether the lag in acquisition from Jones (2024c) is due to deeper study after the initial phase of the study among particularly motivated language learners. At the time of going to press, data collection has commenced with a sample size of approximately N=50, with a pretest, 10 lessons, a post test, and a five-week delayed post-test. Ideally, the study would also be replicated by researchers independent of this project, however, the fact is that replication studies in applied linguistics is the exception rather than the norm.

Finally, it would be useful for the field if the effect of feedback timing on listening were investigated with a larger sample size in order to mitigate effects of attrition. If a larger starting sample size could be recruited, the number of participants with perfect attendance is more likely to be higher, which would result in more granular data. In gaining more granular data, the possibility of observing smaller interactions between variables potentially affecting phonology acquisition increases.

The recommendation of large sample sizes when explaining further research avenues is not a recommendation for typical null hypothesis statistical testing procedures, however, because one of the features of frequentist statistical analysis is that the larger the sample size, the smaller the p values become. A more reliable approach would be to analyse the resultant data using Bayesian methods, possibly alongside frequentist methods. This approach serves the purpose of providing statistics that can be directly compared to Jones (2024e); more people appear to be familiar with frequentist statistics, which more people appear to be familiar with. However, when reporting p values, they should not be reported as significance benchmarks, but simply as the probability that the effect size produced is due to chance.

References

Alptekin, C. (2007). Teaching ELF as a language in its own right: Communication or prescriptivism? *ELT Journal*, *61*(3), 267–268. https://doi.org/10.1093/elt/ccm034

Backhaus, P. (2007). Linguistic landscapes: A comparative study of urban multilingualism in Tokyo. Multilingual Matters.

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Barrios, S. L., & Hayes-Harb, R. (2020). Second language learning of phonological alternations with and without orthographic input: Evidence from the acquisition of a German-like voicing alternation. *Applied Psycholinguistics*, *41*(3), 517–545. https://doi.org/10.1017/S0142716420000077

Bassetti, B. (2023). Effects of orthography on second language phonology: Learning, awareness, perception and production. Routledge.

Bayyurt, Y., & Selvi, A. F. (2021). Language teaching methods and instructional materials in Global Englishes. In A. F. Selvi & B. Yazan (Eds.), *Language teacher education for global Englishes: A practical resource book* (pp. 75–81). Routledge.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, *14*(1), 14–31. https://doi.org/10.1080/10904018.2000.10499033

Carney, N. (2021). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly*, *55*(2), 536–567. https://doi.org/10.1002/tesq.3000

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism: Language and Cognition*, *20*(1), 3–16. https://doi.org/10.1017/S1366728915000863

Cebrian, J., Carlet, A., Gorba, C., & Gavaldà, N. (2019). Perceptual training affects L2 perception but not cross- linguistic similarity. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 929–933.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cheng, J., Matthews, J., Lange, K., & McLean, S. (2022). Aural single-word and aural phrasal verb knowledge and their relationships to L2 listening comprehension. *TESOL Quarterly*, *57*(1), 213–241. https://doi.org/10.1002/tesq.3137

Clark, J. M., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed). Sage.

Cross, J. (2009). Effects of listening strategy instruction on news videotext comprehension. *Language Teaching Research*, *13*(2), 151–176. https://doi.org/10.1177/1362168809103446

Cross, J. (2011). Comprehending news videotexts: The influence of the visual content. *Language Learning & Technology*, *15*(2), 44–68. http://dx.doi.org/10125/44251

Datta, H., Hestvik, A., Vidal, N., Tessel, C., Hisagi, M., Wróblewski, M., & Shafer, V. L. (2020). Automaticity of speech processing in early bilingual adults and children. *Bilingualism: Language and Cognition*, *23*(2), 429–445. https://doi.org/10.1017/S1366728919000099

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

Deterding, D. (1997). The Formants of Monophthong Vowels in Standard Southern British English Pronunciation. *Journal of the International Phonetic Association*, *27*(1–2), 47–55. https://doi.org/10.1017/S0025100300005417

Eales, F., & Oakes, S. (2023). *Speak Out! A2* (3rd ed.). Pearson.

Faerch, C., & Kasper, G. (1980). Processes and strategies in foreign language learning and communication. *Interlanguage Studies Bulletin*, *5*(1), 47–118.

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Field, J. (2011). Into the mind of the academic listener. *Journal of English for Academic Purposes*, *10*(2), 102–112. https://doi.org/10.1016/j.jeap.2011.04.002

Gray, J. (2012). Neoliberalism, celebrity and 'aspirational content' in English language teaching textbooks for the global marke. In D. Block, J. Gray, & M. Holborow (Eds.), *Neoliberalism and applied linguistics* (pp. 86–113). Routledge.

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics*, *45*(1), 34–42. https://doi.org/10.3758/BF03208030

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: A behavioral and ERP study. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00279

Guerrettaz, A. M., & Johnston, B. (2013). Materials in the classroom ecology. *The Modern Language Journal*, *97*(3), 779–796. https://doi.org/10.1111/j.1540-4781.2013.12027.x

Gut, U., Kopečková, R., & Nelson, C. (2023). *Phonetics and phonology in multilingual language development*. Cambridge University Press. http://dx.doi.org/10.1017/9781108992527

Hancock, M., & Donna, S. (2012). *English pronunciation in use. Intermediate: Self-study and classroom use* (Second Edition). Cambridge University Press.

Hardison, D. (2017). Effects of contextual and visual cues on spoken language processing enhancing L2 perceptual salience through focused training. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in Second Language Acquisition*. Routledge.

Hardison, D. M. (1999). Bimodal Speech Perception by Native and Nonnative Speakers of English: Factors Influencing the McGurk Effect. *Language Learning*, *49*(s1), 213–283. https://doi.org/10.1111/0023-8333.49.s1.7

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(4), 579–596. https://doi.org/10.1017/S0142716405050319

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65–94. https://doi.org/10.1177/0267658307071601

Hayes-Harb, R., & Barrios, S. (2021a). Native English speakers and Hindi consonants: From cross-language perception patterns to pronunciation teaching. *Foreign Language Annals*, *n/a*(n/a). https://doi.org/10.1111/flan.12566

Hayes-Harb, R., & Barrios, S. (2021b). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297–326. https://doi.org/10.1017/S0261444820000658

Hertrich, I., Dietrich, S., & Ackermann, H. (2020). The margins of the language network in the Brain. *Frontiers in Communication*, *5*. https://doi.org/10.3389/fcomm.2020.519955

Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N. (2016). Language/Culture Modulates Brain and Gaze Processes in Audiovisual Speech Perception. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep35265

International Phonetic Association (Ed.). (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge University Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. Mouton & Co.

Jamovi team. (2020). *Jamovi* (Version 1.2) [Computer software]. https://www.jamovi.org/

Jarvis, H. (2015). From PPP and CALL/MALL to a praxis of task-based teaching and mobile assisted language use. *TESL-EJ*, *19*(1), 1–9.

Jenkins, J. (2000). The phonology of English as an international language: New models, new norms, new goals. Oxford Univ. Press.

Jones, M. (2017). Alignment of teachers' stated practices with research-based pedagogy in English L2 listening [MA Dissertation]. University of Portsmouth.

Jones, M. (2020). Investigating English Language Teachers' Beliefs and Stated Practices Regarding Bottom-up Processing Instruction for Listening in L2 English. *Journal of Second Language Teaching & Research*, *8*(1), Article 1. http://pops.uclan.ac.uk/index.php/jsltr/article/view/590

Jones, M. (2021). Problems Teaching Listening Online. *The Language Scholar*, *9*, 74–83. https://languagescholar.leeds.ac.uk/problems-teaching-listening-online/

Jones, M. (2024a). Choosing Statistical Analyses for Small Samples. *JAAL in JACET Proceedings*, *6*, 28–32. https://www.jacet.org/JAAL_in_JACET_Proceedings/JAAL_in_JACET_Proceedings_Volume6.pdf

Jones, M. (2024b, May 25). *The reported processes of multilingual listeners*. JALT PanSIG Conference, Fukui University of Technology.

Jones, M. (2024c). Does visual modality improve perceptual acquisition of L2+ English vowels?. [Manuscript submitted for publication].

Jones, M. (2024d). *Listening to Global Englishes: Problems in practice*. [Manuscript submitted for publication].

Jones, M. (2024e). Does visual modality improve perceptual acquisition of L2+ English vowels?. [Manuscript submitted for publication].

Jones, M., & Blume, C. (2022). Accent difference makes no difference to phoneme acquisition. *Teaching English as a Second or Foreign Language Journal--TESL-EJ*, *26*(3), 1–22. https://doi.org/10.55593/ej.26103a3

Joyce, P. (2013). Word Recognition Processing Efficiency as a Component of Second Language Listening. *International Journal of Listening*, *27*(1), 13–24. https://doi.org/10.1080/10904018.2013.732407

Kachru, B. (1988). The sacred cows of English. *English Today*, *4*(4), 3–8. https://doi.org/10.1017/S0266078400000973

Kachru, B. B. (1989). World Englishes and applied linguistics. *Studies in the Linguistic Sciences*, *19*(1), 127–152.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01246

Kaur Phull, D., & Bharadwaja Kumar, G. (2016). Vowel Analysis for Indian English. *Procedia Computer Science*, *93*, 533–538. https://doi.org/10.1016/j.procs.2016.07.264

Keating, P. A., & Huffman, M. K. (1984). Vowel Variation in Japanese. *Phonetica*, *41*(4), 191–207. https://doi.org/10.1159/000261726

Kenanidis, P., Dąbrowska, E., Llompart, M., & Pili-Moss, D. (2023). Can adults learn L2 grammar after prolonged exposure under incidental conditions? *PLOS ONE*, *18*(7), e0288989. https://doi.org/10.1371/journal.pone.0288989

Kiczkowiak, M. (2021). Pronunciation in course books: English as a lingua franca perspective. *ELT Journal*, *75*(1), 55–66. https://doi.org/10.1093/elt/ccaa068

Kikuchi, K., & Browne, C. (2009). English Educational Policy for High Schools in Japan: Ideals vs. Reality. *RELC Journal*, *40*(2), 172–191. https://doi.org/10.1177/0033688209105865

Kim, S. K., & Webb, S. (2022). The Effects of Spaced Practice on Second Language Learning: A Meta-Analysis. *Language Learning*, *72*(1), 269–319. https://doi.org/10.1111/lang.12479

Kobayashi, Y. (2001). The learning of English at academic high schools in Japan: Students caught between exams and internationalisation. *The Language Learning Journal*, *23*(1), 67–72. https://doi.org/10.1080/09571730185200111

Koffi, E. (2021). *Relevant acoustic phonetics of L2 English: Focus on intelligibility.* CRC Press.

Kubota, R. (2015). Inequalities of Englishes, English Speakers and Languages: A Critical Perspective on Pluralist Approaches to English. In R. Tupas (Ed.), *Unequal Englishes: The politics of Englishes today* (pp. 21–41). Palgrave Macmillan.

Kuo, I.-C. (2006). Addressing the issue of teaching English as a lingua franca. *ELT Journal*, *60*(3), 213–221. https://doi.org/10.1093/elt/ccl001

Lambacher, S. (1999). A CALL Tool for Improving Second Language Acquisition of English Consonants by Japanese Learners. *Computer Assisted Language Learning*, *12*(2), 137–156. https://doi.org/10.1076/call.12.2.137.5722

Landry, R., & Bourhis, R. Y. (1997). Linguistic Landscape and Ethnolinguistic Vitality: An Empirical Study. *Journal of Language and Social Psychology*, *16*(1), 23–49. https://doi.org/10.1177/0261927X970161002

Lange, K., & Matthews, J. (2020). Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2 listening comprehension. *Studies in Second Language Learning and Teaching*, *10*(4), 723–749. https://doi.org/10.14746/ssllt.2020.10.4.4

Lindsey, G. (2019). English after RP: Standard British pronunciation today. Palgrave Macmillan.

Lively., S. E. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087. https://doi.org/10.1121/1.410149

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. *From Sound to Sense*, *50*, 381–386.

Lowe, R. J., & Jones, M. (2024). Exploring Duoethnography in ELT Research Ecosystems: Accessibility, Misuse, and Further Horizons. *International Review of Qualitative Research*. https://doi.org/10.1177/19408447241229769

Mackey, A. (2012). Input, interaction, and corrective feedback in L2 learning (1. publ). Oxford Univ. Press.

Martínez-Flor, A., & Usó-Juan, E. (2006). Towards acquiring communicative competence through listening. In E. Usó-Juan & A. Martínez-Flor (Eds.), *Current Trends in the Development and Teaching of the four Language Skills* (pp. 29–46). Mouton de Gruyter. https://doi.org/10.1515/9783110197778.2.29

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, *32*(2), 145–170. https://doi.org/10.1177/0267658315601882

McKay, S. L. (2018). English As an International Language: What It Is and What It Means For Pedagogy. *RELC Journal*, *49*(1), 9–23. https://doi.org/10.1177/0033688217738817

Meierkord, C. (2012). Interactions across Englishes: Linguistic choices in local and international contact situations. Cambridge University Press.

Mora, J. C., & Fullana, N. (2007). Production and perception of English /i/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. *ICPhS XVI*, 1611–1616. https://www.researchgate.net/profile/Joan_C_Mora/publication/242527090_Production_and_perception_of_English_i-I_and_ae-VV_in_a_formal_setting_Investigating_the_effects_of_experience_and_starting_age/links/551bc4db0cf2909047b961f0/Production-and-perception-of-English-i-I-and-ae-VV-in-a-formal-setting-Investigating-the-effects-of-experience-and-starting-age.pdf

Mora, J. C., & Mora-Plaza, I. (2023). From Research in the Lab to Pedagogical Practices in the EFL Classroom: The Case of Task-Based Pronunciation Teaching. *Education Sciences*, *13*(10), Article 10. https://doi.org/10.3390/educsci13101042

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*, *79*(1), 1–43. https://doi.org/10.1515/phon-2022-2018

Munro, M. J. (2021). On the difficulty of defining "difficult" in second-language vowel acquisition. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.639398

Nakatsuhara, F. (2011). *The relationship between test-takers' listening proficiency and their performance on the IELTS Speaking Test* (12; IELTS Research Reports, pp. 1–50). IELTS Australia. http://www.ielts.org/researchers/research/volume_12.aspx

Norouzian, R., de Miranda, M., & Plonsky, L. (2019). A Bayesian Approach to Measuring Evidence in L2 Research: An Empirical Investigation. *The Modern Language Journal*, *103*(1), 248–261. https://doi.org/10.1111/modl.12543

Pawlak, M., & Kruk, M. (2023). *Individual differences in computer assisted language learning research*. Routledge.

Posit Team (2024). *RStudio: Integrated Development Environment for R.* Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Prćić, T. (2012). The role of modernized prescriptivism in teaching pronunciation to EFL university students. In T. Paunović & B. Čubrović (Eds.), *Exploring English phonetics* (pp. 113–124). Cambridge Scholars Publishing.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ravelli, L. J., & van Leeuwen, T. (2018). Modality in the digital age. *Visual Communication*, *17*(3), 277–297. https://doi.org/170.11470/1345707325178217867464436

Roach, P. (2009). *English phonetics and phonology: A practical course* (4. ed., reprinted). Cambridge Univ. Press.

Roberts, R., Buchanan, H., & Pathare, E. (2015). *Navigate (Intermediate B1+)* (Oxford). Oxford University Press.

Rogerson-Revell, P. (2011). English phonology and pronunciation teaching. Continuum.

Rose, H., & Galloway, N. (2019). *Global Englishes for Language Teaching* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316678343

Saito, H. (2020). Acquisition of L2 English intonation by Japanese learners. *Journal of the Institute of Language Research*, *25*, 41–46.

Schrooten, W. (2006). Task-based language teaching and ICT: Developing and assessing interactive multimedia for task-based language teaching. In K. van den Branden (Ed.), *Task-Based Language Education* (pp. 129–150). Cambridge University Press. https://doi.org/10.1017/CBO9780511667282.007

Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching*, *10*(1–4). https://doi.org/10.1515/iral.1972.10.1-4.209

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

Shinohara, Y. (2016). Audiovisual training effects for Japanese children learning English /r/-/l/. *Interspeech 2016.* 204–207. https://doi.org/10.21437/Interspeech.2016-641

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*(4), 1114–1128. https://doi.org/10.2307/1128751

Sparks, R. L., & Ganschow, L. (1991). Foreign Language Learning Differences: Affective or Native Language Aptitude Differences? *He Modern Language Journal*, *75*(1), 3–16. https://doi.org/10.2307/329830

Stefan, A. M., & Schönbrodt, F. D. (2023). Big little lies: A compendium and simulation of p-hacking strategies. *Royal Society Open Science*, *10*(2), 1–30. https://doi.org/10.1098/rsos.220346

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Sydorenko, T., Cárdenas-Claros, M. S., Huntley, E., & Perez, M. M. (2024). Audiovisual input in language learning: Teachers' perspectives. *Language Learning & Technology*, *28*(2), 1–25.

Tsantila, N., & Lopriore, L. (2023). Authenticity in listening. In É. Illés & Y. Bayyurt, *English as a Lingua Franca in the Language Classroom* (1st ed., pp. 147–171). Routledge. https://doi.org/10.4324/9781003258698-9

Van Engen, K. J., Dey, A., Sommers, M. S., & Peelle, J. E. (2022). Audiovisual speech perception: Moving beyond McGurk. *The Journal of the Acoustical Society of America*, *152*(6), 3216–3225. https://doi.org/10.1121/10.0015262

Vettorel, P., & Lopriore, L. (2013). Is there ELF in ELT coursebooks? *Studies in Second Language Learning and Teaching*, *3*(4), 483. https://doi.org/10.14746/ssllt.2013.3.4.3

Vihman, M., & Croft, W. (2007). *Phonological development: Toward a "radical" templatic phonology*. *45*(4), 683–725. https://doi.org/10.1515/LING.2007.021

Walker, R., & Archer, G. (2024). *Teaching English pronunciation for a global world*. Oxford University Press.

Watson, C. I., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change*, *12*(1), 51–68. https://doi.org/10.1017/S0954394500121039

Wolfgramm, C., Suter, N., & Göksel, E. (2016). Examining the role of concentration, vocabulary and self-concept in listening and reading comprehension. *International Journal of Listening*, *30*(1–2), 25–46. https://doi.org/10.1080/10904018.2015.1065746

Yamanaka, S., & Suzuki, K. H. (2020). Japanese education reform towards twenty-first century education. In F. M. Reimers (Ed.), *Audacious education purposes: How governments transform the goals of education systems* (pp. 81–103). Springer International Publishing. https://doi.org/10.1007/978-3-030-41882-3_4

Appendix 1

Jones, M., & Blume, C. (2022). Accent difference makes no difference to phoneme acquisition. *Teaching English as a Second or Foreign Language Journal--TESL-EJ*, *26*(3), 1–22. https://doi.org/10.55593/ej.26103a3

TESL-EJ Teaching English as a Second or Foreign Language
The Electronic Journal for English as a Second Language

Accent Difference Makes No Difference to Phoneme Acquisition

November 2022 – Volume 26, Number 3

https://doi.org/10.55593/ej.26103a3

Marc Jones
Toyo University, Japan
<jones056@toyo.jp>

Carolyn Blume
Technical University Dortmund, Germany
<carolyn.blume@tu-dortmund.de>

Abstract

ELT materials tend to use prestige variety speakers as models, an underlying assumption being that this is needed in order to acquire the phonology necessary to parse English speech (Rose & Galloway, 2019). Global Englishes Language Teaching (GELT) (Galloway & Rose, 2018) provides the potential for movement away from such 'native speaker' ideologies, but lacks empirical evidence. In this study, the use of GELT input in comparison with prestige varieties of English was investigated. Sixteen first-year L1 Japanese university students in an English Medium Instruction programme participated in a self-paced listening study via a learning management system (LMS). All participants were tested on their perception of the English vowels /æ/, /ʌ/, /ɜː/ and/ /ɔː/. After this pretest, they were separated into two groups: using edited TED talks, the experimental group (G) (N=8) watched videos of Global English varieties, and the control group (P) (N=8) watched videos of prestige English varieties. Both groups acquired losses, i.e., immediate posttest scores were mainly lower than pretest scores on vowel identification. Scores were predicted by the variation in interval between lessons and posttest, but not by the varieties of English used. This provides support for the view that GELT is as valid a language teaching approach as using prestige varieties.

Keywords: Global Englishes, listening, multimodality, phonology

Global Englishes and Global Englishes Language Teaching (GELT) (Galloway & Rose, 2017) are an increasingly important part of the English Language Teaching field, gaining prominence as the problematic nature of concepts such as 'native speakerism' (Aneja, 2016; Holliday, 2006; Kiczkowiak, 2019; Lowe & Kiczkowiak, 2019) become subject to growing awareness. Moreover, given the increasing availability of authentic multimodal resources that present a range of English varieties, and learners' exposure to them, GELT potentially promises both a philosophical and pragmatic transformation in English courses of study. However, there is little empirical data concerning how the use of Global Englishes affects language learners' acquisition in measurable terms. Without a better understanding of what linguistic challenges and potential benefits may be associated with such approaches, the development of appropriate pedagogy is not possible.

In this paper, we examine how Global Englishes compare to the use of prestige varieties of English in a self-paced multimodal listening course at a Japanese university in an English Medium Instruction (EMI) course. Stemming from philosophical stances regarding Global Englishes, linguistic models of phonological acquisition, and instructional approaches to multimodal language learning, this contribution provides a rationale for the pedagogical approach employed here while offering empirical analysis of its implementation. While the results do not demonstrate an empirical advantage towards using multimodal Global English materials regarding phonological acquisition, they likewise indicate that there is no marked disadvantage to doing so. In light of the substantial philosophical and conceptual benefits associated with using Global Englishes, we thus argue that these data make a strong case in favour of its utilization.

This paper begins by describing relevant arguments in favour of GELT and extant research that examines attitudes towards prestige and non-prestige varieties of English among English language learners. Highlighting the research gaps in this area regarding acquisition, we will consider empirical data regarding phonological acquisition of prestige and non-prestige varieties. The potential of TED talks as multimodal language learning resources with elements that pertain to Global Englishes will be summarized, based on the literature, before describing the ecological study that was implemented here. Following an explanation of the structure of the learning opportunity and the study design, we will use Bayesian analysis to adequately address the findings in relation to the given sample size. These results will be contextualized in a subsequent discussion before we conclude with an outlook for further areas of inquiry.

Global English Language Teaching: Theory and acquisition

Global English Language Teaching: The theoretical argument(s)

In research literature, definitions of L2 varieties are given in comparison to so-called 'standard varieties,' hereafter referred to as 'standardised varieties' to emphasize our understanding that their use as benchmark varieties is socially constructed. While standardised varieties are generally accorded prestige, other varieties – especially English ones – may be regarded as deficient (Kubota, 2015). This accordance of prestige is linked to the hegemonic status of various nation states, the economic wealth of those states, and also the social capital afforded to citizens of those states or those who may pass as such citizens. Within nation states, further distinctions of prestige accorded to various varieties may also be found. In other words, the prestige attached to language varieties is socially constructed, and depends upon the community in which a given variety is used. This explains in large part why "ranked uppermost among these varieties in terms of prestige and privilege are the two metropolitan Englishes based largely in London and US cities like New York and Washington" (Rubdy, 2015, p. 45). Likewise, extant research examining learners' attitudes towards these so-called non-

standardized Englishes report biases among learners against them, especially as regards accents (e.g. Jacobsen, 2015; Meer, Hartmann & Rumlich, 2021).

In this article we use the term Global Englishes in line with its adoption by Rose & Galloway (2019), who use it as an "inclusive term" (p. 3) to "consolidate research in World Englishes, English as a lingua franca and English as an international language" (p. 4). In short, Global Englishes is a term that refers to use of the language without reference to the problematic notion of nativeness (Holliday, 2006). Given the fact that Global Englishes are largely absent from most commercial ELT materials (Kiczkowiak, 2021), they are also probably underrepresented in practice, leading to a lack of familiarity with such forms that consequently perpetuates discrimination against the speakers of these varieties. This sets up a vicious cycle, ultimately resulting in learners who are ill-prepared for encounters with speakers of non-prestige varieties of English, who view these speakers through a deficit lens (Kubota 2015), and who – depending on their own positionality – likely experience feelings of inefficacy regarding their own competence in their additional language. This has potentially significant ramifications that extend beyond the language learning classroom. As Piller (2016) remarked, "perceiving peripheral speakers as incomprehensible licenses their exclusion from participation in development contexts" (p. 199). In other words, labelling Global Englishes speakers as incomprehensible deskills these speakers and learners, and deprofessionalizes these educators. In English medium instruction (EMI) settings, this attitude sustains a neocolonialist imperative for teachers who are prestige variety speakers and correspondingly marginalises non-prestige speakers of Global Englishes.

Global Englishes accents and phonological acquisition

While advocates of GELT have, for all the aforementioned reasons, been vocal about its theoretical necessity as a corrective to standardised hegemonic language forms, pragmatic issues concerning acquisition have garnered less attention. One of the areas in which varieties are identified is in the phonetic manifestations of their phonology, namely accent. Derwing and Munro (2009) defined speakers' accents as "the ways in which their speech differs from that local variety of English *and* the impact of that difference on speakers and listeners" (p. 479; italics added). In this comparative approach, it is not necessarily the case that a 'foreign accent' is defined in terms of deviation from a prestige accent, but rather in terms of *deviation from what a listener is accustomed to*, or the ease with which a listener can comprehend speech. It is obvious that familiarity contributes to intelligibility. Carey, Mannell and Dunn (2011), for example, investigated ratings by speaking examiners in standardised tests and found that "pronunciation was rated higher by a significant proportion of OPI [oral proficiency interview] examiners when the candidate's interlanguage phonology was familiar and lower when it was unfamiliar" (p. 215, added parentheses). It is therefore clear that exposure to a number of speakers to increase intelligibility is a necessary, but not sufficient, element informing comprehension (Varonis & Gass, 1982). In addition to the linguistic features of grammar and pronunciation, social variables come into play. This includes the racialization of speakers, with a resulting hierarchy of mainly 'white', 'Western', 'native speakers' and an accompanying marginalisation of racialized speakers as 'non-native' (Aneja, 2016). It becomes clear that both speech and image, as well as familiarity, interact to inform perceived intelligibility of oral Global Englishes.

By categorising perceived acoustic phenomena as phonemes, listeners can potentially parse the speech of interlocutors or speech in media. Various models of how this occurs among L2 learners have been developed and tested over decades, with the PAM (L2) model developed by Best and Tyler (2007) most relevant to L2 learners of English living outside L1 English speech communities. The PAM (L2) posits that learners will categorise L2 phones according

to L1 phonemes, and then at a later stage, change this categorisation through creation of categories for L2 phonemes. The speed at which this occurs is dependent upon individual differences, but also relies on sufficient L2 input, with 'sufficient' being defined as a length of time beyond 6-18 months in a L2 environment (Best & Tyler, 2007, p. 21). While research is sparse, it seems that such parameters might hold true for learners to recognize non-standardised varieties as familiar ones.

Multimodal Global English resources and phonological acquisition

Use of available, multimodal Global English resources offers a potential, partial solution to the absence of Global Englishes in commercial materials that contributes to the general lack of familiarity English language learners report regarding these varieties. The popularity of online video platforms such as Youtube and Netflix, as well as TED Talks in particular, as discussed in more detail below, has led to easy access to media in many languages, and also Englishes from many regions. By using easily accessible Global Englishes for language input, the unfamiliarity of varieties underrepresented in ELT can be addressed. Whether this pedagogical approach is sound in terms of phonological acquisition, however, remains largely unexplored. One exception is that Ockey and French (2016) found that self-reported familiarity with British accents correlated with higher comprehension, but did not observe the same for familiarity with Australian accents. Additionally, Saito et al. (2019) found that closeness of listeners' L1 to a L2 speaker's L1 correlated with increased comprehensibility, in keeping with previous findings by Bent and Bradlow (2003). However, vowel acquisition by Japanese L1 users was not investigated in the above studies. Furthermore, the L1 of speakers of the English varieties to be used are all distant from that of our (Japanese) population.

Multimodal phonological acquisition

The language that we use and intend for learners to acquire is multimodal; the auditory information is heard while seeing articulatory gestures (mouth movements) and paralinguistic gestures (other body movements such as pointing). This may aid in phonology acquisition and language learning processes more generally. Ravelli (2018) asserted that "there is a continuum from these inherent forms of multimodality – that is, the visual presence of written language, and the physical presence of spoken language – to the related phenomenon of explicit co-configuration of multiple modes" (p. 434). That is, language events are not subject to a binary of multimodality or single modality but should be considered according to what extent they consist of different modalities. For example, a TED video consists of the speaker's speech, their facial expression and articulatory gestures, whole body movements, audience reaction shots, slide presentations with on-screen text and other modes of communication and therefore may be considered as highly multimodal. When people without sensory impairments or processing issues encounter interlocutors, their attention is focused primarily on the interlocutor's face. This means that visual and linguistic salience "assign a property to a visual/linguistic unit that renders it perceptually more prominent in an array of competing units, which is crucial in cases where selective attention is useful or necessary" (Rácz, 2013, p. 23). Suvorov (2015), for example, found that "L2 test takers spent statistically significantly more time watching content videos (58% of the total time the videos were shown) than context videos (51%)" (p. 476), where content videos are those showing the speaker and their gestures, whereas context videos were those that presented visual information as text or graphics along with speech. This could mean that the participants were focused on the people, including the faces and speech articulation, more often than other visual contexts. While it could also be the case that the content videos in Suvorov's study were simply more interesting than the context videos – a claim that Suvorov himself makes (p.477) – the findings echo those of other interventions that highlight the role of visual reinforcements of articulation processes. Batty (2020) similarly

found, using eye-tracking measures and participant interviews, that learners spent more gazing time on faces than on bodily gestures in video-mediated listening.

In computer-assisted language learning (CALL), multimodality is theorized to be beneficial based on assumptions regarding dual-coding. Mayer (2021) argues that simultaneous processing of visual and auditory information leads to more efficacious encoding. In addition to visual articulatory speech information, paralinguistic features such as gestures, distance and/or physical contact between interlocutors, and prosody further aid comprehension (Hoven, 1999). However, orthographic information may actually interfere with phonological acquisition where L1 and L2 phonology and orthography pairings are incongruent (Hayes Harb & Barrios, 2021; Barrios & Hayes Harb, 2020; Mathieu, 2016), which means closed captions may not be beneficial for this purpose. Therefore, when using multimodal input one must ensure that the different modalities work to facilitate the second language acquisition process rather than interfere. In particular, if there is a potential interlingual interference, caption use should generally be avoided.

Although early uses of CALL for perceptual training (training the ability to perceive differences in speech sounds) reflected theories of learning predicated on repetitive exposure to 'native-like' pronunciation, subsequent technological and conceptual developments have led to an interest in models where learners have greater levels of interaction with technology, as well as greater diversity of speakers in media used (Blake, 2016; Davies et al., 2013; Hubbard, 2017). However, in addition to becoming arguably more inclusive of non-prestige speech varieties, these approaches have tended to focus more on global listening strategies (Hoven, 1999; Jolley & Perez, 2020). Coupled with a potential lack of teacher knowledge about, and skills for, addressing listening skills at the phoneme level (Jones, 2020), research and pedagogy that integrate authentic (here operationalised as materials produced by English speakers, primarily for purposes other than language learning; see Blume, 2022, for more detailed analysis) audiovisual texts for segmental listening skills over a range of speech varieties is sparse.

Multimodal TED talks for listening comprehension of Global Englishes

As in this study, TED (Technology, Entertainment and Design) talks have previously been used in ELT settings with learners of varying L1 backgrounds and level of English knowledge as one source of audiovisual input (Coxhead & Bytheway, 2014; Madarbakus-Ring, 2020; Montero Perez, 2019). TED talks are attractive resources due to the range of topics they address, their easy accessibility, and their effective use of principles of public speaking. They are also clearly multimodal artefacts, consisting of web page, video, options within the video playback interface, spoken text, speaker's gesture and camera angle, as well as the interactive transcript, which allows skipping to the relevant section of the recording when clicked, and/or subtitles. In small scale studies, learners' affective reception of TED talks has been generally high, due to their perceived authenticity (Ahluwalia, 2018; Takaesu, 2017, p. 114) and the range of subjects from which listeners can choose based on their own intrinsic interests (Kozińska, 2021, p. 210; Wu, 2020, p. 33). Additionally, due to the camera angles and production values of TED talks, "Global Englishes speakers are invariably displayed favorably, namely as an expert to be taken seriously regardless of national and/or linguistic backgrounds" (Schildhauer et al., 2021, p. 203). By setting Global English speakers as being of the same status as prestige variety speakers, TED talks can help to legitimize the presence of GELT.

While the library of over 5000 TED talks (https://www.ted.com/) includes speakers of many English varieties, existing descriptions of the talks used in the aforementioned studies do not indicate the varieties with which the studies were carried out. Data collected in these studies reveal that the English learners regularly report that they found the TED talks difficult to

comprehend, at least initially. However, without further descriptions of the specific TED talks utilized by the learners, few conclusions can be drawn about the effect of various varieties on comprehensibility. The reported data suggests that learners found both prestige varieties and varieties from less prestigious and less familiar contexts challenging (Takaesu, 2017; Kozińska, 2021). At the same time, the learners expressed an interest in listening to TED talks with non-prestige varieties, finding both the range of topics and diversity of the speakers motivating (Takaesu, 2017, p. 114; Kozińska, 2021, p. 210).

Although existing studies thus suggest that TED talks may be efficacious for language learning, due to both their multimodality and their affective impact, this prior research is limited in its scope to self-reports of learners' perceived listening facility and attitudinal factors. Nor does it identify the English language varieties that learners are exposed to. Research that empirically examines TED talks, as a source of multimodal input for non-prestige varieties on measures of phonological acquisition, is absent. While a great deal of research has been done in the areas of phonological acquisition and multimodal language learning resources, few ecological studies use authentic materials to evaluate and address learners' phonological acquisition. This gap between theory and practice exists despite the fact that implementing established models of phonology acquisition can be pragmatically realized in multimodal learning opportunities, regardless of the language varieties being targeted.

Methodology

In this paper, we describe a study that examines how Global Englishes compare to the use of prestige varieties of English in a self-paced multimodal listening course. The self-paced course was embedded in an EMI course at a Japanese university. The participants were drawn from the largely L1 Japanese sector of the cohort (approximately 70 percent), with international students making up the remaining 30 percent of each cohort. A convenience sample was taken from the first author's class, based upon informed consent and completion of the self-paced listening course described below. We aim to address the following research questions:

> RQ1: How does GELT affect receptive phonology acquisition of vowels?
>
> RQ2: How does the interval between learning sessions contribute to phonology learning gains?

While the sample size is small, the ecological validity of the intervention is one measure of its validity, with statistical methods selected to accommodate the size of the population. Instead of null hypothesis significance testing using frequentist statistics, Bayes factor analyses were run as more rigorous applications for smaller sample sizes (see Norouzian, de Miranda & Plonsky, 2019). Additionally, generalized linear models (Baayen, Davidson & Bates, 2008, p. 391) with Monte Carlo Markov chain (MCMC) sampling (Thomopoulos, 2013) were used to mitigate power problems typically associated with small samples (for a detailed view see van de Schoot & Miočević, 2020). After analysis, we provide a discussion of the results found and then provide a conclusion along with a comment on the limitations of the study.

Procedures

This study was undertaken with an intact class taught by the first author at a private university in Japan. No learners reported hearing problems or disabilities that affect speech reception or processing. The learners are of approximately equal proficiency level, in that the goal for the end of the first year of their programme of study is to attain an IELTS score of 6.0 or above. This is roughly intermediate level, or deemed sufficient by higher education institutions for entry to an exchange programme. Learners took a pretest and then were separated into two groups paired roughly by score, i.e. two learners gaining the same or almost the same score

were assigned to different groups to create balance across conditions. From the intact class, sixteen learners returned consent forms permitting their data to be used for this study, therefore the sample size was N=16, of which eight students were in each condition described below. One learner reported Chinese L1 but Japanese as a dominant language, and was placed in the experimental condition group described below. The learners gained academic credit for completion and partial completion of the self-paced listening course described below. No financial rewards were offered or given.

Tests

Learners listened to 32 CVC words recorded by the first author using a Blue Snowball iCE microphone and PRAAT software (Boersma & Weenink, 1992), eight each of the vowels /æ/, /ʌ/, /ɔ:/, and /ɜ:/, labelled in plots and tables as TRAP, STRUT, THOUGH and NURSE respectively. (See the full word list in Appendix 3). The reason for focusing on vowels is that Japanese does not have as many central vowels as English (Keating & Huffman, 1984) and therefore during listening, the inability to accurately perceive or discriminate these English vowels may cause comprehension difficulties. Additionally, the selected vowels have similar formant ('undertone') values: /æ/, /ʌ/, and /ɜ:/ have similar first and second formant values, and all vowels have similar third formant values according to Deterding (1997, p. 49), meaning that while classification for L1 users may be straightforward, the differences between them may not be salient enough for easy L2 acquisition. Furthermore, given the breadth of work on consonant acquisition (for example Sheldon & Strange, 1982; Lambacher, 1999; Lotto, Sato & Diehl, 2004; Sueyoshi & Hardison, 2005) in comparison to the scant work on vowel acquisition for L1 Japanese learners of L2+ English, (Ingram & Park, 1997; Nishi & Kewley-Port 2005, 2008), more work is warranted.

The test was the same for both pretest and posttest conditions, with different question sequences in the pretest and posttests. The didactic materials and the tests were delivered through a learning management system (Moodle, 2020), and the vowels identified by choosing a picture of a common noun that shares the same vowel as the recorded word. The pictures provided were of the words *cat, sun, door,* and *bird* for the vowels /æ/, /ʌ/, /ɔ:/, and /ɜ:/ respectively.

Self-paced listening course

The course content was provided through the lesson module in Moodle. Each lesson consisted of four sets of a vowel identification exercise in the context of a short word (CVC, CVCC or CCVC) with automated feedback (Figure 1), then an utterance decoding task containing the word from the vowel identification (Figure 2). This provided not only vowel discrimination as a discrete activity but integrated it into a realistic decoding task, providing affordances for participants to integrate discrimination into their decoding and parsing processes. Participants then watched a 10-minute excerpt of the TED talk and subsequently provided a summary. Captions were not available due to the assumption of phonological interference given the incongruence between English orthographic and Japanese L1 phonological systems as discussed previously. This structure was repeated over five lessons, and was intended to be completed on a weekly basis, for a total of five weeks. Group P listened to videos of speakers of standardised prestige English varieties, while Group G listened to videos of speakers of Global Englishes, detailed in Table 1, below.

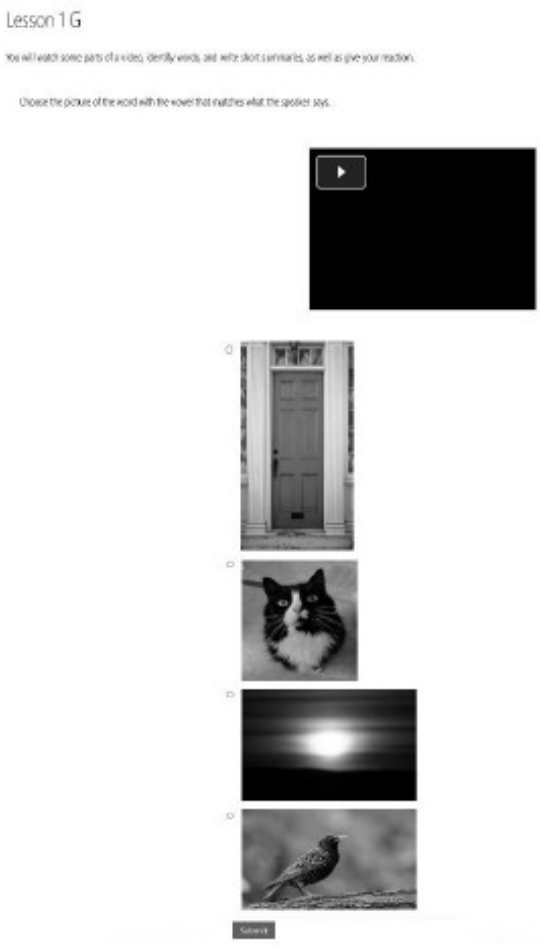


Figure 1 Screenshot of vowel identification task

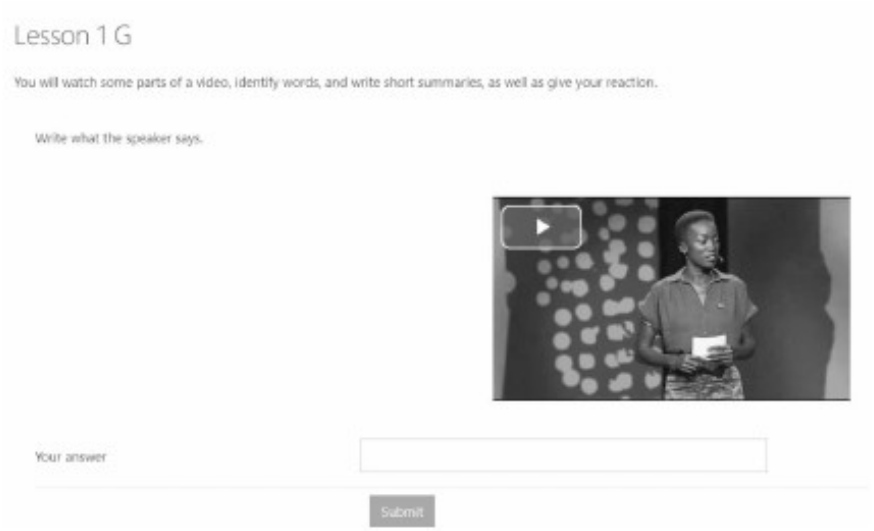


Figure 2 Screenshot of utterance decoding task

Table 1. Table of TED talks used for each lesson. Group G - Global Englishes; Group N – Prestige 'Native' Englishes. Variety of English or additional language influence on English given after year. Varieties only defined by L1 influence and not dialect or specific region.

Theme	Group G Talk	Group N Talk
Urbanization	Adegbeye (2017) Nigerian	Speck (2013) American
Maternal healthcare	Hegde (2021) Indian	Howell (2018) American
Environmentalism	Jun (2021) Chinese	Francis (2019) English
Personal finance	Belle (2021) Kenyan	Gibson (2019) American
Political disagreement	Cekic (2018) Turkish/Danish	Pearlman (2019) American

Results

Table 2. Mean pretest and posttest scores, time (in days) from L1-L2, L3-L4, L4-L5, L5-post, and mean standard deviation of intervals.

Condition	Mean Pre	Mean Post	Mean Interval L1_2 (Days)	Mean Interval L2_3 (Days)	Mean Interval L3_4 (Days)	Mean Interval L4_5 (Days)	Mean Interval L5_Post (Days)	Mean Interval SD
G	26.5	24.50	16.37	2.31	2.72	6.86	4.83	688.62
N	27.5	24.37	4.38	6.49	1.93	4.99	9.29	170.50

Pretest and posttest scores were used to calculate gains for each participant in R (R Core Team, 2020), both overall and for each vowel studied. These were then analysed in a Bayesian Welch's t-test using BEST (Kruschke & Meredith, 2021) for two-sample comparison, i.e., cross-group comparison. Beyond this, GLMs were constructed in Stan (Guo et al., 2020) and Rstanarm (Gabry & Goodrich, 2020) with MCMC sampling in order to find what factors contributed to the resultant scores. These models were then compared using the bridgesampling (Gronau et al., 2020) and BEST (Kruschke & Meredith, 2021) packages. Bayesian priors were set at null (i.e. default broad priors) due to the lack of similar studies and available prior evidence to provide a suitable prediction.

As can be seen in Table 2, there were negative gains for both groups P and G on average, with a minority of learners making minimal positive gains. Gains in group G were generally seen through greater acquisition of TRAP while gains in group P were made in acquisition of NURSE (Table 3). However, using the guidelines for interpreting Bayes factors in Norouzian, de Miranda and Plonsky (2019, p. 252) both these gains are marginal and may be attributed to chance. This is reinforced by the Bayes factors for all the gains.

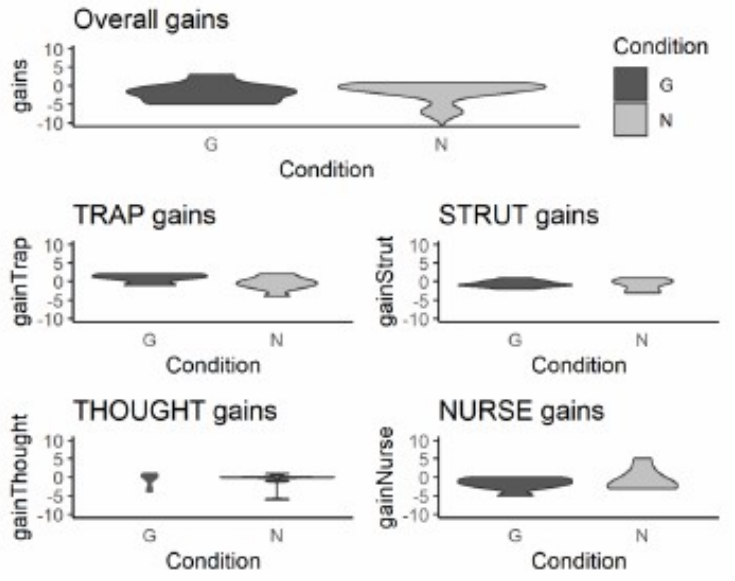


Figure 3. Mean gains by group overall and by vowel

Table 3. t-tests for group G vs. group P

Description	t Score	Bayes Factor
Overall gains	0.50	0.32
TRAP gains	1.90	0.22
STRUT gains	0.39	0.77
THOUGHT gains	0.24	0.77
NURSE gains	-0.95	3.08

When analysing the GLMs to account for the trend in gains, it appears that condition is irrelevant, as seen in Table 4. The increase in Bayesian factor between 'Gains, condition, participant and interval standard deviation: no interaction vs interaction' and 'Gains, participant and interval standard deviation: no interaction vs interaction' shows that the largest interaction is due to the standard deviation in time intervals between lessons and posttest. Due to analysis of linear models in comparison with one another, ceiling effects are unlikely to be present, as can be observed in the series of similar linear models being compared.

Table 4. Bayes factors for the comparisons of different GLMs

Description	Bayes Factor
Gains, condition and intervals no interaction vs interaction	Inf
Gains, condition, participant and interval standard deviation: no interaction vs interaction	0.45
Gains, participant and interval standard deviation: no interaction vs interaction	118.91
Gains, condition, participant and individual vowel gains: no interaction vs interaction	0.00
TRAP gains, condition, participant and interval standard deviation: no interaction vs interaction	0.002
STRUT gains, condition, participant and interval standard deviation: no interaction vs interaction	0.64
THOUGHT gains, condition, participant and interval standard deviation: no interaction vs interaction	1.10
NURSE gains, condition, participant and interval standard deviation: no interaction vs interaction	0.18

Discussion

As can be seen from the results of both the MCMC sampled GLM and the Bayesian t-test, according to Nourozian et al.'s (2019) guidelines for Bayes factor interpretations (p. 252), results of the t-tests show that there is insufficient evidence to support the use of prestige varieties of English over Global English, with relatively low t scores and Bayes factors very close to suggesting only anecdotal evidence strength. Therefore, we can state that there is effectively no difference in phonology acquisition between groups that listened to prestige variety Englishes and non-prestige Global Englishes. The only factor that seemed to affect phonology acquisition was the standard deviation of intervals between lessons in the self-paced course. The evidence thus suggests that GELT is no less effective an approach than teaching using prestige models in a self-paced listening course based on TED talks. The learners using Global English materials did not have significant differences or worse losses than the prestige variety group. While this may seem intuitively or experientially obvious to practitioners using GELT in their day-to-day teaching, there is little so far in the extant literature to provide empirical evidence for the benefits of GELT. This study, despite its relatively small scale, thus offers data to support a GELT approach, in that it is not inferior to the use of prestige models.

In addition to marginalizing both teachers and learners who do not reflect stereotypes of so-called 'native speakers,' a prescriptivist, prestige-based standard holds as an ideal a largely unattainable goal that is at odds with contemporary notions of communicative competence. Rather, given patterns in global deterritorialisation (see Jacquemet, 2005, p. 261), the ability to communicate transculturally is of utmost importance. Of course, whether using prestige or non-prestige varieties, transcultural communication is possible. Our point here is that, due to prestige varieties having already been widely used in ELT, parsing of prestige varieties poses fewer problems for learners whereas less prestigious Global Englishes have been absent from materials and this lack of familiarity may cause problems. Ensuring that learners are equipped to communicate among an increasingly mobile global population means that an understanding of English as it is used as a 'lingua franca' (ELF) (Jenkins, 2000; Seidlhofer, 2011) thus needs to be both a philosophical and a didactic priority. (In this context, we use the term Global Englishes to refer to what speakers say, whereas ELF refers to the use of English as a common language among speakers of different languages.) While the dominance of English in transnational academic, popular, and economic contexts admittedly perpetuates a problematic neoliberal agenda (Jacobsen, 2015; Kubota, 2015, 2021), fostering learners' ability to function in this context offers them a degree of agency. That is, by teaching not only with prestige varieties but also non-prestige varieties, learners gain skills to communicate more effectively across a wider array of communities using English.

In light of the fact that the absence of regular intervals is associated with the lack of phonological acquisition, more research is needed to examine the role played by students' proclivity to complete all the activities at once, as revealed by completion data obtained via the learning management system. It is plausible that progressing in irregular increments throughout the course of study might inform the lack of gains, regardless of Global English variety. In keeping with research regarding practice for L2 vocabulary retention (Bloom & Shuell, 1981) and L2 syntax (Bird, 2010), regular attention over a period of time might be more likely to contribute to measurable gains, though this is yet to be confirmed for phonology. If this is the case, it would be important to address the pedagogical issues involved in managing student progress in a self-paced course in particular contexts.

Alternatively, more intensive or more prolonged learning opportunities might be necessary to facilitate gains in vowel perception, although this is currently unclear. As suggested by studies about familiarity with accents more generally (cf. Varonis & Gass 1982) and discussed above, it is unclear what amount of exposure is necessary to improve phonological acquisition. Due to the easier measurement of pronunciation, this appears to frequently be a proxy measure of perceptual phonology. Tyler & Best (2007) suggested that research populations of language learners tend not to differ greatly in their phonology after 18 months living in the community where they study, although their focus, to be clear, is on pronunciation. Mora and Fullana (2007) found that length of study does not affect Catalan/Spanish bilinguals' perception of English /æ/-/ʌ/ or/i/-/ɪ/ contrasts (pp. 1615-1616). As summed up by Flege and Bohn (2020), "[i]t is unknown at present how much L2 input is needed to form phonetic categories in an L2 and optimally adapt them to everyday use. This may depend, at least in part, on the uniformity of the L2 speech input that is received" (p.13). In short, the time required for L2 vowel acquisition may be subject to individual differences, such as time spent within different languages, but existing research is scant and contradictory. More research in this area would be useful to all speech learning researchers, whether related to English or other languages.

The question also remains whether these particular videos, or videos in general, are unsuitable media for phonological acquisition. McCrocklin (2012) demonstrated that there was no significant effect for video over audio for /i/ - /ɪ/ contrasts in an ESL class in the USA with a large number of L1 Chinese speakers. However, Navarra and Soto-Faraco (2007) found that

video was more beneficial for perceiving /e/-/ɛ/ contrasts than audio alone. Furthermore, Becker and Sturm (2015), while not reaching statistical significance in comparing audio and visual modes for L2 French listening comprehension, found that video had a larger effect size than audio only. Other factors that might hypothetically affect the suitability of the media could, instead, lie in the complexity or unfamiliarity of the topics or length of the video excerpts, which ran to approximately 10 minutes, with very short excerpts preceding the final full 10-minute viewing. Unfamiliarity with listening tasks without the affordance of negotiating meaning with the speaker may also be possible factors contributing to our results.

We believed that using videos for listening comprehension would facilitate acquisition through greater visual salience. One reason for this may be that language processing seems to be linked to the articulatory features. That is, when visual articulatory information is perceived, it is processed in the same area of the brain as auditory information, Broca's region (Glanz et al., 2018). This means that if auditory input can be made more visually salient, or that learners have more than one source of evidence for the phonetic information they are attempting to parse, they are more likely to be successful in their efforts. This appears to be the case in studies by Hardison (2003; Sueyoshi and Hardison, 2005), where consonants were acquired more readily by learners exposed to audiovisual information than those exposed to audio-only information. Inceoglu (2022), on the other hand, found mixed results of an audiovisual effect in the perception of L2 French vowels, outcomes being mediated by prior word knowledge and individual vowel features. This is potentially due to the nasality of some French vowels in contrast to the absence of nasal vowels in English and, which may contribute to a lack of visual salience among the former. Like Inceoglu, our results suggest that the videos do not significantly contribute to acquisition. As we did not have an audio-only condition, direct comparisons are difficult. However, it may be that audiovisual effects vary for consonants and vowels, regardless of language.

It appears likely that viewing multiple videos shortly before the end of their course of study could be one reason for the losses in perceptual discrimination of the target vowels. This could be attributed to working memory fatigue or waning selective attention. It may also be the case that this loss is part of a developmental trajectory, similar to the 'U-shaped' acquisition trajectory theorised by Shirai in relation to lexical learning (1990). Whether the processes and theories postulated regarding 'U-shaped' trajectories for initial language acquisition and second language semantic items are relevant for phonological acquisition has not been established.

Limitations

One of the clearest limitations of this study is the small sample size, and while this is somewhat mitigated by the use of Bayesian statistical approaches and MCMC sampling of the data in creating the GLM, the effect sizes reported should be approached with caution. While ecologically valid, the study is only one intervention with a relatively small scope. More work is required to account for the lack of phonology gains among participants. While other factors might play a role, it seems that the self-paced nature of the course led to some students 'cramming' the practice into unevenly spaced, and brief intervals, to the detriment of their phonology acquisition. This behaviour might reflect the timing of the sequence at the end of the academic year, fatigue from multiple competing obligations due to the close of the semester, or the unique situation still informed by COVID-19-related restrictions and extensive digital interactions. Such 'limitations' are, in the context of a study embedded in an intact language learning setting, not to be considered flaws in the research design. Rather, they highlight the human factors that play into formal language learning in a classroom setting that cannot ever be eliminated and should therefore not be excluded from empirical studies.

At the same time, ecological validity in one context does not necessarily result in generalisability across other contexts. Therefore, we welcome more research by scholars working in different contexts to build on the methodology in this study. This would develop a broader evidence base for phonology acquisition studies in pedagogical rather than laboratory-based work. Additionally, it would be highly useful for the field of ELT in general if further investigations into the use of GELT for second language acquisition were undertaken. GELT is becoming more important in ELT as the effects of globalisation and the exposure to less hegemonically dominant English varieties becomes a necessity rather than a curiosity limited to the university classroom.

Conclusion

While many questions are raised from the project, we would like to believe that we have contributed somewhat in showing that GELT is not a lesser approach to English teaching than the prestige-variety status quo. In short, teachers should not be reluctant to use a range of English varieties in their classroom practice.

While the outcomes of this study can be considered positive as regards GELT, they are sobering in light of the lack of gains made regarding receptive phonology among both groups. According to the data, the lack of gains can be attributed to the lack of regular intervals between undertaking lessons for many of the learners. While adult learners may be expected to be able to make appropriate decisions about their own learning, the patterns of interaction with the material in the LMS suggest that this is not universally the case, at least among this cohort. Given the effect on learning outcomes, as evidenced by this admittedly small study, more attention needs to be given to balancing the pedagogical need for effective monitoring with young adult learners' autonomy-related needs. The data showing how pacing interacts with acquisition clearly illustrate the ways in which second language acquisition research and didactic design need to go hand-in-hand.

Regardless of these caveats, it is clear that phonology acquisition is necessary for listening, because acoustic information needs to be parsed in order to be meaningful, and that this can demonstrably occur through input provided by speakers of any variety of English. Indeed, it seems apparent that a GELT approach (Galloway & Rose, 2014; Rose & Galloway, 2019) would be more appropriate than the status quo, not only for purposes of increasing language acquisition, but also for the purposes of reducing implicit colonialist, 'native speaker' ideologies in teaching. As Derwing and Munro (2014, p. 219) pointed out in reference to communication between L1 users and conversational participants with alternative L1s, the responsibility for intelligible communication resides with both interlocutors. At the same time, they made it clear that "[b]ecause the activation of biases is a phenomenon that takes place within the listener, addressing its effects requires changes in the listener rather than the speaker. Just as one cannot legitimately expect the target of a racist act to correct the problem by changing color, one cannot expect a non-native speaker to 'drop' an accent because others respond negatively to it. Rather, ample evidence indicates that accented speech is a normal aspect of second language development, and there is no evidence that accents can be routinely eliminated through pedagogical intervention" (2014, p. 221). As such, we disagree with points of view such as that of Tschirner (2011), who stated that, in pedagogical designs, "speakers or language that is difficult to understand" should be excluded (p. 35), since this begs the question of how these speakers or the language could ever become easily understood if they are, or it is, perpetually excluded. Seen in this light, addressing learners' perception of Global English pronunciations must combine efficacious phonological training with pedagogical measures to address potentially discriminatory attitudes.

About the Authors

Marc Jones is a language teacher and researcher at Toyo University. His research interests are primarily listening and phonology, extending to various other aspects of language teaching. ORCID ID: 0000-0002-2004-1809

Carolyn Blume holds the chair for digitally-mediated teaching and learning in the Dortmund Competence Center for Teacher Education & Research (DoKoLL) at the Technical University Dortmund, Germany. A member by courtesy of the Faculty of Cultural Studies (English), Dr. Blume's research focuses on initial and ongoing teacher education for English language teaching, specifically in the areas of inclusion and diversity, and digital media and digitality. Dr. Blume is the 2022 recipient of the IDEA Award, honoring university level teaching that is inclusive and that prepares teachers for diversity in their classrooms. ORCID ID: 0000-0002-4788-593X

Acknowledgements

Support for this article was provided in part by DoProfil, part of the "*Qualitätsoffensive Lehrerbildung*," a joint initiative of the Federal Government and the *Länder* which aims to improve the quality of teacher training. The programme is funded by the Federal Ministry of Education and Research. The authors are responsible for the content of this publication.

To Cite this Article

Jones, M. & Blume, C. (2022). Accent Difference Makes No Difference to Phoneme Acquisition. *Teaching English as a Second Language Electronic Journal (TESL-EJ), 26* (3). https://doi.org/10.55593/ej.26103a3

References

Ahluwalia, G. (2018). Students' perceptions on the use of TED Talks for English language learning. *Language in India, 18(*12), 80-86.

Aneja, G. A. (2016). Rethinking nativeness: Toward a dynamic paradigm of (non)native speakering. *Critical Inquiry in Language Studies*, *13*(4), 351–379. https://doi.org/10.1080/15427587.2016.1185373

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barrios, S. L., & Hayes-Harb, R. (2020). Second language learning of phonological alternations with and without orthographic input: Evidence from the acquisition of a German-like voicing alternation. *Applied Psycholinguistics*, *41*(3), 517–545. https://doi.org/10.1017/S0142716420000077

Batty, A. O. (2020). An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing*, 026553222095150. https://journals.sagepub.com/doi/10.1177/0265532220951504

Becker, S., & Sturm, J. L. (2017). Effects of audiovisual media on L2 listening comprehension: A preliminary study in French. *CALICO Journal*, *34*(2), 147–177. https://doi.org/10.1558/cj.26754

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in second language speech learning: In honor of James Emil Flege* (Vol. 17, pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*(4), 635–650. https://doi.org/10.1017/S0142716410000172

Blake, R. (2016). Technology and the four skills. *Language Learning & Technology, 20(*2), 129–142. http://dx.doi.org/10125/44465

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*(4), 245–248. https://doi.org/10.1080/00220671.1981.10885317

Blume, C. (2022). Authentizität. In S.J. Schierholz (Ed.), *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)* (Online). Mouton de Gruyter.

Boersma, P., & Weenink, D. (1992). *PRAAT: Doing phonetics by computer*. (6.1.02) [Computer software]. http://www.praat.org/

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*(3), 643–664. https://doi.org/10.1017/S0142716411000518

Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, *28*(2), 201–219. https://journals.sagepub.com/doi/10.1177/0265532210393704

Coxhead A., & Bytheway, J. (2014) You will not blink: learning vocabulary using online resources. In D. Nunan & J.C. Richards (Eds.), *Language learning beyond the classroom* (pp. 65–74). Routledge.

Davies, G., Otto, S. E. & Rüschoff, B. (2013). Historical perspectives on CALL. In M. Thomas, H. Reinders & M. Warschauer (Eds.), *Contemporary computer-assisted language learning* (pp. 19–38). Bloomsbury.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490. https://doi.org/10.1017/S026144480800551X

Derwing, T. M. & Munro, M. J. (2014). Training native speakers to listen to L2 speech. In J. Levis & A. Moyer (Eds.), S*ocial dynamics in second language accent* (pp. 219–238). de Gruyter Mouton.

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, *27*(1–2), 47–55. https://doi.org/10.1017/S0025100300005417

Flege, J., & Bohn, O. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Gabry, J., & Goodrich, B. (2020). *rstanarm: Bayesian Applied Regression Modeling via Stan.* https://CRAN.R-project.org/package=rstanarm

Galloway, N., & Rose, H. (2014). Using listening journals to raise awareness of global Englishes in ELT. ELT Journal, 68(4), 386–396. https://doi.org/10.1093/elt/ccu021

Galloway, N., & Rose, H. (2018). Incorporating Global Englishes into the ELT classroom. *ELT Journal, 72*(1), 3–14. https://doi.org/10.1093/elt/ccx010

Glanz, O., Derix, J., Kaur, R., Schulze-Bonhage, A., Auer, P., Aertsen, A., & Ball, T. (2018). Real-life speech production and perception have a shared premotor-cortical substrate. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-26801-x

Gronau, Q. F., & Singmann, H. (2021). *Bridgesampling: Bridge sampling for marginal likelihoods and Bayes factors*. https://github.com/quentingronau/bridgesampling

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297–326. https://doi.org/10.1017/S0261444820000658

Holliday, A. (2006). Native-speakerism. ELT Journal, 60(4), 385–387. https://doi.org/10.1093/elt/ccl030

Hoven, D. (1999). A model for listening and viewing comprehension in multimedia environments. *Language Learning & Technology*, *3*(1), 73–90. http://dx.doi.org/10125/25058

Hubbard, P. (2017). Technologies for teaching and learning L2 listening. In C. A. Chapelle & S. Sauro (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (pp. 93–107). John Wiley & Sons.

Inceoglu, S. (2021). Language experience and subjective word familiarity on the multimodal perception of non-native speakers' vowels. *Language & Speech*, 1. https://journals.sagepub.com/doi/10.1177/0023830921998723

Ingram, J. C., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*(3), 343–370. https://doi.org/10.1006/jpho.1997.0048

Jacobsen, U. C. (2015). Cosmopolitan sensitivities, vulnerability, and global Englishes. *Language and Intercultural Communication*, *15*(4), 459–474. https://doi.org/10.1080/14708477.2015.1031674

Jacquemet, M. (2005). Transidiomatic practices: Language and power in the age of globalization. *Language & Communication*, *25*(3), 257–277. https://doi.org/10.1016/j.langcom.2005.05.001

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford University Press.

Jolley, K. & Perez, M. (2020). Exploring marginalised communities with online student portfolios using Google Drive and TEDx Talks. In K.-M. Frederiksen, S. Larsen, L. Bradley & S. Thouësny (Vorsitz), CALL for widening participation: Short papers from EUROCALL 2020. (pp. 143-148)

Jones, M. (2020). Investigating English language teachers' beliefs and stated practices regarding bottom-up processing instruction for listening in L2 English. *Journal of*

Second Language Teaching & Research, *8*(1), 52–71. http://pops.uclan.ac.uk/index.php/jsltr/article/view/590

Keating, P. A., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*, *41*(4), 191–207. https://doi.org/10.1159/000261726

Kiczkowiak, M. (2019). Students', Teachers' and recruiters' perception of teaching effectiveness and the importance of nativeness in ELT. *Journal of Second Language Teaching & Research*, *7*(1), 1–25. https://pops.uclan.ac.uk/index.php/jsltr/article/view/578

Kiczkowiak, M. (2021). Pronunciation in course books: English as a lingua franca perspective. *ELT Journal*, *75*(1), 55–66. https://doi.org/10.1093/elt/ccaa068

Kozińska, K. (2021). TED talks as resources for the development of listening, speaking and interaction skills in teaching EFL to university students. *Neofilolog, 56*(2), 201–221. https://doi.org/10.14746/n.2021.56.2.4

Kruschke, J., & Meredith, M. (2021). *BEST: Bayesian Estimation Supersedes the t-Test.* https://CRAN.R-project.org/package=BEST

Kubota, R. (2015). Inequalities of Englishes, English speakers and languages: A critical perspective on pluralist approaches to English. In R. Tupas (Ed.), *Unequal Englishes: The politics of Englishes today* (pp. 21–41). Palgrave Macmillan.

Kubota, R. (2021). Global Englishes and teaching culture. In A. F. Selvi & B. Yazan (Eds.), *Language teacher education for global Englishes: A practical resource book* (pp. 135–143). Routledge.

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, *12*(2), 137–156. https://doi.org/10.1076/call.12.2.137.5722

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. *From Sound to Sense*, *50*, 381–386.

Lowe, R. J., & Kiczkowiak, M. (2016). Native-speakerism and the complexity of personal experience: A duoethnographic study. *Cogent Education*, *3*(1), 1264171. https://doi.org/10.1080/2331186X.2016.1264171

Madarbakus-Ring, N. (2020). Developing graded TED Talks to integrate academic vocabulary into listening lessons for pre-sessional learners. In *The TESOL encyclopedia of English language teaching* (pp. 1–7). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118784235.eelt0990

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, *32*(2), 145–170. https://doi.org/10.1177/0267658315601882

Mayer, R. E. (2021). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/9781108894333

McCrocklin, S. (2012). Effect of audio vs. video on aural discrimination of vowels. *TESL-EJ*, *16*(2), 1–16. https://www.tesl-ej.org/wordpress/issues/volume16/ej62/ej62a2/

Meer, P., Hartmann, J., & Rumlich, D. (2021). Attitudes of German high school students toward different varieties of English. *Applied Linguistics*, *amab046*. https://doi.org/10.1093/applin/amab046

Montero Perez, M. (2019). Technology-enhanced listening: How does it look and what can we expect? In N. Arnold & L. Ducate (Eds.), *Engaging language learners through CALL: From theory and research to informed practice* (pp. 141–178). Equinox.

Montero Perez, M., van den Noortgate, W. & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Moodle. (2020). *Moodle*. https://moodle.com/

Mora, J. C., & Fullana, N. (2007). Production and perception of English /i/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. *ICPhS XVI*, 1611–1616. https://tinyurl.com/Mora-J-2007

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*(1), 4–12. https://doi.org/10.1007/s00426-005-0031-5

Nishi, K., & Kewley-Port, D. (2005). Training Japanese listeners to identify American English vowels. *The Journal of the Acoustical Society of America*, *117*(4), 2401–2401. https://doi.org/10.1121/1.4786064

Nishi, K., & Kewley-Port, D. (2008). Non-native speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research: JSLHR*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Norouzian, R., de Miranda, M., & Plonsky, L. (2019). A Bayesian approach to measuring evidence in L2 research: An empirical investigation. *The Modern Language Journal*. https://doi.org/10.1111/modl.12543

Ockey, G. J., & French, R. (2016). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics*, *37*(5), 693–715. https://doi.org/10.1093/applin/amu060

Piller, I. (2016). *Linguistic diversity and social justice: An introduction to applied sociolinguistics*. Oxford University Press.

Rácz, P. (2013). *Salience in sociolinguistics: A quantitative approach*. De Gruyter Mouton.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.0.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ravelli, L. J. (2018). Multimodal English. In P. Seargeant, A. Hewings, & S. Pihlaja (Eds.), *The Routledge handbook of English language studies* (pp. 434–446). Routledge.

Rose, H., & Galloway, N. (2019). *Global Englishes for language teaching*. Cambridge University Press. https://doi.org/10.1017/9781316678343

Rubdy, R. (2015). Unequal Englishes: The native speaker, and decolonization in TESOL. In R. Tupas (Ed.), *Unequal Englishes: The politics of Englishes today* (pp. 42–58). Palgrave Macmillan.

Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do second language listeners perceive the comprehensibility of foreign-accented speech?: Roles of first language profiles, second language proficiency, age, experience, familiarity, and metacognition. *Studies in Second Language Acquisition*, *41*(5), 1133–1149. https://doi.org/10.1017/S0272263119000226

Schildhauer, P., Zehne, C., & Schulte, M. (2021). Encountering Global Englishes in the ELT classroom through audio-visual texts: The example of TED talks. In M. Callies, S. Hehner, & P. Meer (Eds.), *Glocalising teaching English as an international language: New perspectives for teaching and teacher education in Germany* (pp. 198-214). Routledge.

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

Shirai, Y. (1990). U-shaped behavior in L2 acquisition. In H. Burmeister & L. P. Rounds (Eds.), *Variability in second language acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum, Vol. 2* (pp. 685–700).

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing*, *32*(4), 463–483. https://doi.org/10.1177/0265532214562099

Takaesu, A. (2017). TED Talks as an extensive listening resource for EAP students. In K. Kimura & J. Middlecamp (Eds.), *Asian-focused ELT research and practice: Voices from the far edge* (pp. 108–120). IDP Education.

Thomopoulos, N. T. (2013). *Essentials of Monte Carlo simulation*. Springer. https://doi.org/10.1007/978-1-4614-6022-0

Trudgill, P., & Hannah, J. (2017). *International English: A guide to varieties of English around the world* (6th edition). Routledge, Taylor & Francis Group.

Tschirner, E. (2011). Video clips, input processing and language learning. In W. M. Chan, K.N. Chin, M. Nagami, & T. Suthiwan (Eds.), *Media in foreign language teaching and learning* (pp. 25–42). De Gruyter.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, *4*(2), 114–136. https://doi.org/10.1017/S027226310000437X

Wu, C.-P. (2020). Implementing TED Talks as authentic videos to improve Taiwanese students' listening comprehension in English language learning. *Arab World English Journal*, 6, 24–37. https://doi.org/10.24093/awej/call6.2

Appendix 1: List of TED Talks used

Adegbeye, O. (2017). *Who belongs in a city?* https://www.ted.com/talks/olutimehin_adegbeye_who_belongs_in_a_city

Belle, R. A. (2021) *The emotions behind your money habits*.

https://www.ted.com/talks/robert_a_belle_the_emotions_behind_your_money_habits?language=en

Cekic, Ö. (2018). *Why I have coffee with people who send me hate mail.* https://www.ted.com/talks/ozlem_cekic_why_i_have_coffee_with_people_who_send_me_hate_mail

Francis, A. (2019). *How to get everyone to care about a green economy.* https://www.ted.com/talks/angela_francis_how_to_get_everyone_to_care_about_a_green_economy

Gibson, E. (2019). *The true cost of financial dependence.* https://www.ted.com/talks/estelle_gibson_the_true_cost_of_financial_dependence

Hegde, A. (2021). *The life-saving tech helping mothers make healthy decisions.* https://www.ted.com/talks/aparna_hegde_the_life_saving_tech_helping_mothers_make_healthy_decisions

Howell, E. (2018). How we can improve maternal healthcare -- before, during and after pregnancy. https://www.ted.com/talks/elizabeth_howell_how_we_can_improve_maternal_healthcare_before_during_and_after_pregnancy

Jun, M. (2021). An interactive map to track (and end) pollution in China. https://www.ted.com/talks/ma_jun_an_interactive_map_to_track_and_end_pollution_in_china

Pearlman, E. (2019). How to lead a conversation between people who disagree. https://www.ted.com/talks/eve_pearlman_how_to_lead_a_conversation_between_people_who_disagree

Speck, J. (2013). The walkable city. https://www.ted.com/talks/jeff_speck_the_walkable_city

Appendix 2: List of R packages used

Gabry, J., & Goodrich, B. (2020). *rstanarm: Bayesian Applied Regression Modeling via Stan.* https://CRAN.R-project.org/package=rstanarm

Gronau, Q. F., & Singmann, H. (2021). *bridgesampling: Bridge Sampling for Marginal Likelihoods and Bayes Factors.* https://github.com/quentingronau/bridgesampling

Guo, J., Gabry, J., Goodrich, B., & Weber, S. (2020). *rstan: R Interface to Stan.* https://mc-stan.org/rstan/

Kruschke, J., & Meredith, M. (2021). *BEST: Bayesian Estimation Supersedes the t-Test.* https://CRAN.R-project.org/package=BEST

Simonsohn, U., & Gruson, H. (2021). *groundhog: The Simplest Solution to Version-Control for CRAN Packages.* https://CRAN.R-project.org/package=groundhog

Spinu, V., Grolemund, G., & Wickham, H. (2021). *lubridate: Make Dealing with Dates a Little Easier.* https://CRAN.R-project.org/package=lubridate

Wickham, H. (2021). *tidyverse: Easily Install and Load the Tidyverse.* https://CRAN.R-project.org/package=tidyverse

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2.* https://wilkelab.org/cowplot/

Appendix 3: List of words tested

All recordings are available to download from the project's Open Science Framework page. https://osf.io/xb9ce/?view_only=e420caa1444a42aba486a129e71920a0

bad, ban, board, born, bun, burn,

call, can, cork, cup

dad

fawn, fun

girl

ham, heard, hug, hut,

kern,

lash, luck

nurse

pause, pearl, pork

sad, shut, sword,

torn, turn

yearn

zap

Appendix 2

Jones, M. (2024). Choosing Statistical Analyses for Small Samples. *JAAL in JACET Proceedings*, *6*, 28–32. https://www.jacet.org/JAAL_in_JACET_Proceedings/JAAL_in_JACET_Proceedings_Volume6.pdf

Choosing Statistical Analyses for Small Samples

JONES, Marc*

*Toyo University
jones056@toyo.jp

Abstract
Small samples are frequent in language teaching research due to participant recruitment and attrition problems. Additionally, some types of quantitative methods require sample sizes beyond the size of a usual intact class, making frequentist analysis of work with single classes problematic due to underpowered analyses. However, if little or no research exists that answers a particular research question, is it logical to prolong the existence of such as gap in the literature, or should research be carried out with small samples? This presentation provides a deeper rationale for the analysis techniques used in an already existing study (Jones & Blume, 2022) which had a very small number of participants, but which necessitated a quantitative analysis due to the logic of the research questions. The rationale for a use of Bayesian methods is discussed, along with limitations of Bayesian analysis, such as researcher and reviewer unfamiliarity. Furthermore, use of different linear regressions such as ANOVA, GLM and GLMM are discussed with regard to controlling for individual differences and environmental factors in studies.

Keywords: Bayesian analysis, GLMM, null hypothesis statistical testing

1. Introduction

Applied linguistics has many studies with small samples, perhaps because researchers are less well funded than those in the hard sciences and therefore cannot always afford to compensate participants at the level of, for example, clinical trials. However, small samples can lead to underpowered research, particularly with frequentist statistics. There are several choices facing researchers regarding he analysis of data in quantitative studies (and the quantitative data in mixed methods studies). In this article, the choices made in regard to the analyses in Jones and Blume (2022) are examined.

2. Small samples

One of the main problems in research in applied linguistics may be participant recruitment, especially when one considers that commonly recommended sample sizes for t-tests are N>32 and for chi-squared tests are N>50. These are not the types of sample that correspond to typical practical language classes, and therefore recruiting from one's own students can still result in coming up short. Several issues related to participant recruitment and attrition are noted in McKinley and Rose (2017). There are several contributing factors surrounding recruitment difficulty. If recruitment of volunteer participants from outside one's own classes is necessary, the fact is that rapport has not been established and therefore there is no trust-based relationship. This means that informed consent is not straightforward for potential participants because they do not know whether the researcher is likely to keep the promises made in the information sheet accompanying any consent forms, particularly in the case when ethics board approval is not required by the institution. In addition, students are frequently busy with other classes when researchers have the time to conduct their studies, and lunch times are also far from ideal. There is also the fact that if researchers ask students to give up their time, they are potentially giving up the potential to earn money in part-time work. In sum, there is usually no inherent benefit to students in participating in research unless there is some kind of outcome that they desire, such as, for example, test practice for an upcoming standardised test.

The problems of recruitment listed above are compounded by those of participant attrition. In-person attendance problems can also contribute to participant attrition, because illness can be as difficult to predict as it is to prevent. Inconvenience for participants is also a contributing factor. Participants may agree to research, providing 'informed consent' without actually paying sufficient attention to what is required of them. Furthermore other unforeseen problems may lead to attrition.

However, consideration needs to be paid to whether research should be carried out at all with small samples if no research exists on the topic. Questions researchers may need to ask themselves are whether there is a possibility that a reasonable conclusion can be reached. If not, can comparison be made between groups or samples? Can data be estimated or simulated the data? If none of these questions can be answered with affirmatively, is recruitment of further participants reasonable in the time available? Is a suitable analysis for the data possible?

Plonsky and Oswald (2014) assert that "quantitative L2 research produces substantially larger effects than those in many other fields" (p. 890) but this could be due to L2 research's small samples, large effects required with them, and publication "bias toward studies with statistically significant findings" (Plonsky & Oswald, 2014, p. 891). One

way of avoiding overly small samples in applied linguistics research is by using power analysis tools such as G Power (Faul et al., 2009). This software ensures that a sample size for sufficient detection of given minimum effect size is known prior to commencing data collection and therefore, as mentioned above, whether further recruitment is required after the first round of data collection.

Below, the reasons for existing problems with small samples due to frequentist statistics are explained and alternatives in the form of Bayesian analyses provided.

3. Misuse of frequentist statistical analysis

Null hypothesis statistical testing (NHST) is the process of analysing data to test whether the researchers' hypothesis (often referred to, somewhat confusingly for those new to quantitative research, the *alternative hypothesis*) or the *null hypothesis* (the hypothesis that the intervention has no effect) is supported. One danger here is that hypothesis 'support' is different to a hypothesis being 'correct'. When researchers' alternative hypotheses are supported by their analyses, bold claims are sometimes made on the basis of a value that is arbitrarily selected, as explained in this section. NHST is the norm in applied linguistics research yet there are issues with them relating to both power and also the way that interpretation of statistical values are interpreted.

The way that p values tend to be used is as a measure of significance, in that they should be lower than a chosen alpha. Alpha can be simplified here as the acceptable probability of a false positive result, but it actually means the acceptable probability that a more extreme test statistic in the direction of the alternative hypothesis may be observed if the null hypothesis is true. A p value of <= 0.05 being a typically accepted value for significance in language teaching research. NHST is problematic because p values decrease as the sample size increases. Thus, it is viable for research projects with sufficient resources to recruit volunteers to participate in a study until an acceptable p value for significance is achieved, which is known as p-hacking. Trafimow et al., (2018) argue that despite using ever smaller p values in the psychological sciences, that this does not solve the problem of the use of an alpha for NHST, and that decisions about evidence supporting the theoretical accuracy of a hypothesis should be made over a number of studies carried out by different independent groups of researchers instead of on the basis of a single study.

Alternatively, applied linguistics researchers can diverge from the status quo and forgo NHST. The dangers of NHST are outlined in Trafimow (2024) in that alpha values are entirely arbitrary (although it is possible to make meaningful decisions regarding alpha levels regarding sample size). If one is conducting analyses using frequentist statistics, knowing what the p value represents is important. It is not simply 'significance' but the probability that the result found is due to chance. Therefore, 0.05 is perhaps too high for medical research NHST, but is conversely too low for an exploratory analysis of a small sample. That is, that a is drug found to be safe having a 5 per cent probability of being unsafe is obviously far too high, but restricting research on a teaching intervention's efficacy to a 5% chance of a false positive test seems overly cautious. Although calculating the sample required to detect a desired effect size at the p-level of one's choice is possible in GPower (Faul et al., 2009), researchers need to actually reflect upon these values, and in particular, if they use frequentist statistics, to consider the use of p values and apply common sense to the size of their chosen alpha.

Furthermore, due to the binary nature of NHST, studies are either 'successful' in confirming or negating hypotheses. Such an approach is misguided when studies are novel or are investigating cutting-edge research questions. The position put forward in this article is that NHST has no place unless exploratory studies have already been conducted. Given that p values are used to provide a benchmark value for confirming null hypotheses, it can be stated that instead of using p values as a decision trigger parameter, a nuanced view should be taken, particularly when considering effect and sample sizes. However, Bayes Factor Analysis (BFA) may be even more effective in exploratory analysis, because a probability index of how well data fits one of two models can be obtained.

4. Bayesian analysis

Bayesian data analysis is philosophically different to the frequentist statistical analysis used to conduct NHST. Frequentist statistical analysis assumes that the probability of an occurrence is based upon frequency. Conversely, Bayesian statistics is not based upon simple frequency but upon prior evidence collected to support one model in comparison to another. Previous data collection (even in other studies) can inform Bayesian priors, a coefficient that informs the model comparison (Zondervan-Zwijnenburg et al., 2017). Mackey and Ross (2015) advocate use of Bayesian techniques, which they state are "optional when researchers are testing hypotheses predicated on grounded theoretical arguments, and in the present illustration, for testing framework-driven operationalizations of proficiency" (p. 329). However, they do not mention the strengths of Bayesian analysis for exploratory work in the early stages of theory development, which, arguably constitutes a large amount of SLA and speech learning work.

5. Case study: Jones and Blume (2022)

In this section the data analysis decisions taken in Jones and Blume (2022) are described from an emic point of view. The article in question describes the quantitative part of a study exploring whether there is a reason to maintain the status quo in English language teaching of using the prestige varieties of English in listening for perceptual phonology acquisition. By the term 'prestige varieties' we are describing the prestige that has been traditionally accorded to them, essentially the varieties prevalent in the colonial mercantile cities of predominantly Anglophone societies (Rubdy, 2015). In contrast, the set of inputs we used as an alternative were by speakers of Global Englishes, the varieties of English used outside the countries in Kachru's (1984) inner circle, and frequently by non-white people.

The rationale behind the study was that while there is

rhetoric behind use of prestige varieties as the 'default' listening input, there is a shortage of empirical work. Additionally, prior work on L2+ much of the vowel acquisition with Japanese learners of English has been laboratory based (Nishi & Kewley-Port, 2007, 2008), rather than using students in intact classes, thus potentially lacking ecological validity.

The study was undertaken with a grouped pretest-posttest design, which allowed for gains (or in fact, losses) in perceptual ability to be measured easily. Learners of English with L1/dominant language Japanese (N=16) undertook a pretest of the vowels the vowels /æ/, /ʌ/, /ɔ:/ and /ɜ:/, after which they were paired by score and each pair was split, placing them into opposing groups. Both groups took 5 online lessons using TED talks, training the vowels /æ/, /ʌ/, /ɔ:/ and /ɜ:/ in consonant-vowel-consonant word contexts, the words in utterance contexts, then the utterances in the context of a video edited for length. One group, (n=8) received input from prestige varieties, and the other (n=8) received input from Global Englishes speakers. Due to the exploratory nature of the study, not only was the effect overall investigated but also the effect at the individual vowel level.

5.1 Data analysis decisions

The study in Jones and Blume (2022) was designed to investigate whether there was an effect in use of either prestige variety (i.e. so-called 'native speaker' varieties of English) on comparison to Global Englishes varieties (Galloway, 2013) for perceptual vowel acquisition. The study was designed in such a way that gains in perceptual ability could be measured and analysed in a straightforward way. The intuitive decision was to use paired t-tests, because this is frequently the orthodox way to compare the performances or developments of two matched groups in different conditions. The unorthodox decision to use t-tests rather than the Wilcoxon Signed Rank test or Mann-Whitney U-test was due to the difference in statistical power affordances. While the latter two tests assume non-normal distribution, and a t-test assumes normal distribution, violation of these assumptions is possible when the means behave as normal, and because the t-tests were conducted with Bayesian analysis using MCMC sampling, the simulated sampling can be assumed to be sufficiently robust. Additionally, where a Pearson p-value tends to be taken with t-tests, Wilcoxon and Mann-Whitney tests, Bayes factor analysis allows for a view of how well the data fit the model.

The models in the t-tests were actually that prestige varieties resulted in higher gains than Global Englishes (PV > GE) and that prestige varieties did not result in higher gains than Global Englishes (PV ≤ GE), and these tests were undertaken for all four vowels tested overall, and for the four individual vowels tested. Attentive readers may notice that this is similar to NHST, however, the difference is the purpose and use of benchmarking values. The Bayes factor is not used in the same way that p values are used in NHST, but are indicators of the probability that the model fits the data.

As stated above, the study had a total sample of N=16, with two groups, each of n=8, therefore, while it was possible to test this as a hypothesis, because the central limit theorem does not hold in Bayesian theory, it would be unwise due to this being exploratory work to investigate the use of PV and GE in an ecologically valid setting, i.e. outside of a laboratory setting. Additionally, warnings on Bayesian hypothesis testing are given in Schad et al. (2023). The question regarding data analysis was to what extent the evidence supports either model. The analysis showed no support for PV > GE, therefore we concluded that no serious support could be given for the theory that prestige varieties are better models for perceptual vowel acquisition from audiovisual media.

Beyond whether data analysis of the pretest and posttest, understanding the factors contributing to gains (or losses) in perceptual acquisition were required. The most commonly applied linear regressions to analyse such factors are analysis of variance (ANOVA) and analysis of covariance (ANCOVA). However, more detail of interactions can be taken using a Generalised Linear Mixed Model (GLMM), which use Monte-Carlo Markov Chain (MCMC) sampling of data.

GLMM affords scrutiny of the data and the interactions of the variables. In using MCMC to sample the data and using the variables in these samples to create models and compare these models' interactions, contributing factors to the behaviour of variables in the data set can be observed. Calculating BF, allows researchers to understand the probability of the model fitting to the data. As the p value is merely an expression of the probability of gaining a more extreme statistical score under a null hypothesis, the BF provides researchers greater confidence in the factor's relationship to the data as a whole rather than as a probability of comparison of two means reaching an arbitrary benchmark. This is not to say that Bayesian GLMM are a panacea for research; there is a learning curve in the use of Bayesian analyses in general, and GLMM generally require the ability to use programming languages such as R, Python or Julia for complex calculations. In spite of the difficulties, the depth of analysis with meaningful results is of greater benefit to researchers than the ease of using preset analyses without considering their suitability. Above all, the best analysis for the data at hand is required, and in order to make the decision about what is best, alternative analyses also need to be considered.

The main reason for choosing to proceed on the work for Jones and Blume (2022) was that there was a lack of empirical work on the use of GEs for L2+ acquisition and also a lack of work in a classroom setting on vowel learning. Pilot studies had also been conducted at the tail end of the COVID pandemic conditions, resulting in sample attrition and extremely noisy data, although the opportunity for fine-tuning processes was possible. However, because the participants were my own students, as part of an intact class, I understood from the outset that steps were needed to mitigate any problems that were likely to occur. This being the case, proceeding with a methodology for small samples was the plan from the outset, and qualitative data was also collected to allow for triangulation. Despite the advanced planning, the study still had room for failure, such as students becoming tired or ill and becoming unwilling or unable to complete the

work required to allow for useful findings. However, if frequentist statistics had been the only resource at my disposal, I believe that the likely outcome of the study would have been failure due to lack of statistical power and the lack of confidence in making inferences from descriptive statistics alone.

5.2 Alternative analyses

The reasons for using t-tests with paired samples is one of orthodoxy, however, I violated one of the rules in that t-tests assume data is normally distributed. Our data was very much not normal. However, given the uniform lack of normality across all of the groups in the different t-tests, it was assumed that this would not be a significant problem. Additionally, because the t-tests were being used for exploratory purposes rather than for NHST, and the use of Bayes factors to gauge how strong the evidence was for different models meant that a nuanced interpretation of the data analysis was possible, i.e. there was no binary true/false hypothesis testing, but instead a consideration of whether an existing supposition not yet tested empirically held up to scrutiny.

However, alternative non-parametric tests are available, such as the Wilcoxon ranked sign test Wilcoxon (1945), which is also available for Bayes factor analysis (Barch & Chechile, 2023) and the Mann-Whitney U test (Mann & Whitney, 1947). The downside of these non-parametric tests is that they have lower statistical power, and therefore with small sample sizes there is a greater chance that evidential strength goes undetected.

As mentioned above, ANOVA and ANCOVA are popular ways of analysing factors contributing to learning gains. Other alternatives include the ordinary Generalised Linear Model (GLM). The GLM does not provide quite as much detail as GLMM, because the GLM is an analysis of interactions between factors without random effects. Additionally, a common linear regression is also possible, but again lacks the granular detail of the GLMM.

For a frequentist alternative to MCMC, bootstrapping (Larson-Hall & Herrington, 2010) helps create robust statistical power in a similar way to Markov-chains, which resamples smaller samples to increase power. As an additional option, complex analyses could be foregone in favour of using descriptive statistics such as mean differences, standard deviation differences, and confidence intervals (Plonsky, 2015).

6. Conclusion

Jones and Blume (2022) used Bayesian analyses because as a methodology, Bayesian analysis lends itself to exploratory work more than frequentist work does, which more often than not is aimed at NHST. While it is entirely possible to interpret p values in a nuanced way, the tendency is to use them as an arbitrary value to give binary yes/no answers. Bayes factors provide an alternative to this, although Norouzian et al. (2019) caution that their guidelines for Bayes Factors should not be interpreted in similar ways to the use of p values. In conclusion, the use of any tool for overly simplified decision making needs to be avoided and peer reviewers need to be vigilant and advise researchers on good research practices in data analysis, as they would do for other aspects of research methodology.

Notes

The original data and analysis script for R is available at OSF https://osf.io/xb9ce/. This means that the script may be built upon and further, multifaceted analyses of the data are possible.

References

Barch D. H., & Chechile R. A. (2023). *DFBA: Distribution-Free Bayesian Analysis*. R package version 0.1.0, https://CRAN.R-project.org/package=DFBA

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Jones, M., & Blume, C. (2022). Accent difference makes no difference to phoneme acquisition. *TESL-EJ, 26*(3), 1–22. https://doi.org/10.55593/ej.26103a3

Larson-Hall, J., & Herrington, R. (2010). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics, 31*(3), 368–390. https://doi.org/10.1093/applin/amp038

Mackey, B., & Ross, S. J. (2015). Bayesian informative hypothesis testing. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 329–345). Routledge.

McKinley, J., & Rose, H. (Eds.). (2017). *Doing research in applied linguistics: Realities, dilemmas, and solutions*. Routledge.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496–1509. https://doi.org/10.1044/1092-4388(2007/103)

Nishi, K., & Kewley-Port, D. (2008). Non-native speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research: JSLHR*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Norouzian, R., de Miranda, M., & Plonsky, L. (2019). A Bayesian approach to measuring evidence in L2 research: An empirical investigation. *The Modern Language Journal, 103*(1), 248–261. https://doi.org/10.1111/modl.12543

Plonsky, L. (2015). Statistical power, p values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23–45). Routledge.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Rubdy, R. (2015). Unequal Englishes: The native speaker, and decolonization in TESOL. In R. Tupas (Ed.), *Unequal Englishes: The politics of Englishes today* (pp. 42–58). Palgrave Macmillan.

Schad, D. J., Nicenboim, B., & Vasishth, S. (2023). *Data aggregation can lead to biased inferences in Bayesian linear mixed models and Bayesian ANOVA: A simulation study* (arXiv:2203.02361). arXiv. https://doi.org/10.48550/arXiv.2203.02361

Trafimow, D. (2024). *Methodological issues in psychology: Concept, method, and measurement*. Routledge.

Trafimow, D., … Marmolejo-Ramos, F. (2018). Manipulating the alpha level cannot cure significance testing. *Frontiers in*

Psychology, *9*. https://doi.org/10.3389/fpsyg.2018.00699
Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & Schoot, R. V. de. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development, 14*(4), 305–320. https://doi.org/10.1080/15427609.2017.1370966

Appendix 3

Jones, M. (2024c). *Does visual modality improve perceptual acquisition of L2+ English vowels?* [Manuscript submitted for publication].

Does visual modality improve perceptual acquisition of L2+ English vowels?

Abstract

Findings from neuroscience shows that the prefrontal cortex of the brain can be stimulated by speech, both acoustically and visually, when seeing a speaker articulate speech sounds. This suggests that as speech is a multimodal phenomenon, then multimodal input may provide richer cueing to facilitate acquisition of additional languages. However, there are few studies which have investigated input modality, and fewer still in classroom settings. The current study investigated the acquisition of L2+ English vowels by L1 Japanese learners in an intact class (N=32) at a university in Tokyo. Half the class were exposed to audiovisual input and the other half audio only for six weeks. Immediate posttest gains were insufficiently different between group, but of the ten learners that participated in a delayed posttest, gains increased in the audiovisual group. Although the evidence is weak, it provides the basis for further investigation.

Keywords: English, modality, phonology, vowels

The ways modality affects second or additional language phonology acquisition is so far largely unknown. Few studies have been conducted using naturalistic audiovisual stimuli to investigate the effects of seeing speakers' articulatory gestures on phoneme acquisition, and this is particularly the case regarding vowels. Furthermore, there is a lack of work taking place in classroom settings, even when participants are recruited from student bodies.

Laboratory-based SLA studies have shown that orthography can have both positive and negative effects (see Hayes-Harb & Barrios, 2021 for a review). Outside of orthography, gated video has been used to facilitate consonant acquisition (Hardison, 1999, 2003, 2017) and wave forms can facilitate phonological acquisition (Motohashi-Saigo & Hardison, 2009). However, in a classroom setting, these methods are not always practical due to the extensive preparation and/or technological expertise required.

In this article, the literature on speech learning and multimodality in speech perception is reviewed, along with literature on the potential effects of raising visual salience for phonological acquisition, and which is then assessed for how it may apply to vowel acquisition in classroom-based contexts. In the methodology section, the details of vowel identification tests and the instruction procedures which provide ecological validity for the study are provided. The results, are provided, with Bayes factor analyses, after which a conclusion is drawn referring to the results in relation to the extant literature. Finally, the findings and their ramifications are discussed, with recommendations for further research as well as the current study's limitations.

1 Literature review

1.1 Speech learning

Practical considerations of how L2+ phonemes are acquired in classroom practice and how the process can be facilitated more efficiently is largely absent from the literature. One reason may be that laboratory studies avoid the messiness of classroom conditions. Unfortunately, in forgoing classroom conditions there may be an equal consequence of forgoing ecological validity. In the absence of classroom-oriented L2+ speech perception research, L2+ listening remains difficult in spite of the ever-expanding sources of internet audiovisual media available inexpensively or for free. That is not to say that there is a complete absence of classroom-oriented research (two key examples are Saito et al., 2019; and Mora and Mora-Plaza, 2023). However, a lot of classroom-oriented phonology research is focused upon production measures (i.e. speaking), which, as far as it concerns perceptual abilities, is only a proxy, and in fact learners have larger phonologies to apply to L2+ perception than to L2+ production. In this section both laboratory and classroom studies investigating speech learning are examined in order to consider how segmental phonology acquisition can be facilitated effectively.

1.1.1 PAM-L2 + Vocab model

Best's (1995) The Perceptual Assimilation Model (PAM) proposed that naive listeners of a language categorised speech sounds in one of the following ways:

- same as L1 category,
- good fit to L1 category,
- poor fit to L1 category,
- uncategorized according to L1.

  Speech sounds that are further away from L1 categories were likely to be uncategorized according to L1. However, this leads to paying no mind to the sounds, and further learning is not modelled directly with the PAM. An extension to the PAM was developed, PAM-L2 (Best & Tyler, 2007), because "the phonological level is central to the perception of L2 speech by (second language) learners, who are developing an L2 (or interlanguage) system, in a way that it cannot be for L2-naïve listeners perceiving unfamiliar nonnative speech" (p. 23). With the development of the L2 interlanguage system, it is proposed that categorization is unnecessary, as perceptions can be used at the basic gestural level, phonetic level, up to the abstracted phonological level (Best & Tyler, 2007, p. 25). The PAM-L2 assumes that there is a unified perceptual phonology across languages, and therefore if L2 phonemes can map to L1, no new entries are required. It is when novel sounds are presented that category assignment related to L1 categories is required (Tyler, 2019).

Further work related to vowel sounds related to the PAM-L2 is seen in work on the Vocab Model (Bundgaard-Nielsen, et al., 2011, 2012), itself an extension to the PAM-L2, where vocabulary size is seen as a factor positively influencing L2 phonology perception. The Vocab Model posits that phonological and vocabulary acquisition development may affect one another, i.e. any effect is bidirectional and that "changes in perception *initially* precede those in production. This may also be reflected in the observed lag between developmental change in L2 perception and production" (Bundgaard-Nielsen et al., 2012, p. 660; emphasis

in original). The model applies to classroom instruction and classroom-centred research in that the development of different language systems (e.g. phonology, lexis) can affect one another. In other words, the interactions between acquisition in the different language systems may account for the U-shaped acquisition pattern posited by Shirai (1990). In short, according to the PAM(-L2) and Vocab models, new categories for L2 phones may not be necessary but categorisation as L1 phonemes is likely, and the development of vocabulary acquisition potentially interacts with phonological acquisition.

1.1.2 SLM-r

Flege (1995) devised the Speech Learning Model to account for the way that long-term L2 users develop their L2 phonology. The main way in which the SLM differs from the Best's (1995) PAM is that the PAM was devised to account for the ways in which naive users or non-users of a language perceive its phonology whereas the SLM is for more accomplished, long-term users of the language (Best & Tyler 2007). Additionally, the PAM accounts for perceptual abilities, whereas the SLM accounts for both perception and production, due to perception being a prerequisite for production. However, neither the PAM nor the SLM were devised to account for instructed language learning. In the SLM revised (SLM-r) (Bohn & Flege, 2018) a theoretical grounding was given to opposing the Critical Period Hypothesis (CPH) (e.g. Lenneberg, 1967), stating that L1 speech learning mechanisms stay open to learners of L2, giving credence to frequency-based theories of speech learning and language acquisition. In doing so, the experiences of language teachers can be linked to theory: while very fluent second/additional language speakers may be relatively rare among white English speakers, for much of the world, high proficiency in more than one language is normal (see Bokamba, 2018, for examples from people of the African diaspora; Hakuta et al., 2003, discusses US Census data and language proficiency). An interesting aspect of the SLM is that Flege (1995) states that the phonological categories of L1 and L2 blur together as the L2 develops; that is, not only does the L1 affect L2 but L2 development affects L1 categorisation. It should also be noted here that Flege is discussing adult learners, not children, and he is discussing acquisition of phonology, not syntax or lexis, and not discussing L2 users through deficit-based perspective.

1.1.3 Second Language Linguistic Perception (L2LP) model

Escudero's (2005) Second Language Linguistic Perception (L2LP) model considers speech learning as an interaction between L1 and L2 phonology development. The model assumes that "L2 learners use the auditory information that differentiates sound categories differently from native listeners" (Escudero, 2005, p.85). The model involves the following parts:

1. optimal L1 perception and optimal target L2 perception;
2. the L2 initial state;
3. The L2 learning task;
4. L2 development;
5. the L2 end state.

   (Escudero, 2005, pp. 87-113)

The PAM-L2 posits that existing L1 phonological representations inform the phonemes that listeners hear: the items heard are categorised according to how well they fit the L1 categories, but this becomes more like an L2 map after enough input has been turned into uptake to facilitate some degree of L2 categorisation (Best & Tyler, 2007). Additionally, the SLM states that L1 and L2 categories blur with input (Flege, 1995). However, Escudero (2005) states that "a separation of perceptual mappings from sound representations leads to an adequate comparison of the perception systems involved" in the L2LP (p. 86). However the L2LP initial state is explained by PAM (Best, 1995) and SLM, where L1 categories dictate the L2 categories, but as L2 development progresses, particularly with regard to the L2 learning task and L2 development, where noticing (Schmidt, 1990) should occur for the learner to develop their phonological learning. This way of considering learners' phonological perceptions according to different developmental stages of perceptual phonological acquisition allows prediction of what can and cannot likely be perceived to be more fine grained. However, learners may differ in *how* they map L2 sounds to L1 categories (Mayr & Escudero, 2010) and therefore researchers may consider pretesting

and/or gathering preliminary data from teachers, results from standardised tests or from the learners themselves to be necessary.

1.1.4 Natural Referent Vowel framework (NRV)

The Natural Referent Vowel framework (NRV) (Polka & Bohn, 2011) refers to a concept where vowels at the periphery of the vowel space act as mental references for the rest of the vowels in an individual's phonology. Essentially, in the case of English, the vowels /i/, /u/, /ɑ/ and the pair of American and British equivalents /a/ (American) and /æ/ (British) act as natural cardinal vowels, as opposed to the primary cardinal vowels used as comparison references by phoneticians (e.g. International Phonetic Association, 1999, p. 12). These vowels act as reference points to which the other vowel phonemes need to be sufficiently distanced from in order to be appropriately perceived (or produced, but this is beyond the scope of the current article). In doing so, the vowels that function as natural referents allow the provision of a heuristic of acquisition difficulty. For a given L2 vowel, theoretically the difficulty is relative to the distance from a NRV and L1 vowels which are not equivalent.

Flege and Mackay (2004) found that perception of vowels by relatively early onset and late onset L1 Italian learners of English had perception differences. Late learners had higher similarity to age-matched L1 English users' perception of vowel contrasts compared to early onset learners, L1 Italian students who were compared to age matched L1 English students for /ɛ-æ/, /ɒ-ʌ/, or /i-ɪ/ contrasts (Flege & Mackay, 2004, p. 20). The presence of vowels which are on the periphery of the vowel space, namely /æ/, /ɒ/ and /i/, suggest the potential of NRV effects. That is, while peripheral vowels themselves may be universals, vowels in relatively close proximity to them may be less easily learnt than more central vowels if those central vowels have no 'neighbours'.

Additionally, when contrasting vowels, Polka and Bohn (2011) found that a contrast from more peripheral to more central was easier to discriminate than more central to more peripheral. Furthermore, while the bias toward NRV can be overcome, the bias may re-emerge in processing speed or when listening under adverse conditions, such as in noise. This bias may be useful to bear in mind for testing, in order to ensure that contrasts in both

directions are present in test items in order to provide a thorough assessment, and also to understand that what occurs in a test (arguably a condition where some learners feel a pressure to perform) is only a proxy for real-world listening.

1.2 Multimodality in speech perception

Findings in neuroscience by Glanz et al. (2018) suggest that the prefrontal cortex of the brain, in particular Broca's area, is activated by not only hearing speech but also seeing speech articulated. While Glanz and colleagues' findings relate to first language, Callan et al. (2014) found that second-language auditory input also activates Broca's area, and Berezutskaya et al. (2020) believe that the prefrontal cortex "simply elicits a more basic, fundamental response to perceived speech" (p. 4604). In short, there is activity in Broca's area with speech perception, and it has been noted to be associated with speech motor activation. Furthermore, it is activated by both first and second language speech. Thus, there is evidence of speech perception being a multimodal phenomenon, and therefore multimodal input should aid the acquisition of L2+ phonology, and in turn affect L2+ listening development.

1.2.1 Orthography

Orthography appears to be an obvious way to increase visual salience. However, English is not a highly orthographically transparent language: grapheme-phoneme correspondences do not map one-to-one in all languages, and therefore orthography is difficult to harness in order to increase visual salience of phonological items (Hayes-Harb & Barrios, 2021). Therefore, while use of orthography can work between some pairs of languages (Showalter and Hayes-Harb, 2013) it is unsuccessful for others (Barrios & Hayes-Harb, 2020), and because English is orthographically less transparent, L1 Japanese learners are likely to experience no advantage in using it to aid phonology acquisition. Furthermore, Bassetti (2023) found that orthography interferes with phonology acquisition and "results in the learning of a sound that is neither attested in the target phonological system nor present in the auditory input." (pp. 81-2) and that errors in acquired L2+ phonology are resistant to

instruction. These findings add further evidence to support avoiding phonology for L1 Japanese learners of English, at least in the early stages of acquisition.

1.2.2 High Variability Phonetic Training (HVPT)

High Variabilty Phonetic Training (HVPT) is a type of drill training using short stimuli, usually CV or CVC words, to facilitate phonetic perception and/or production. The method was first carried out by Jamieson and Morosan (1986) to study acquisition of English L2 dental fricatives by L1 French Canadian learners. However, acquisition of vowels has also been examined (e.g. Mora & Fullana, 2007; Mora et al., 2022; Nishi & Kewley-Port, 2007, 2008). One problem with some HVPT studies are that labels may rely upon orthography, which could affect uptake, particularly across languages sharing the same orthographic system, or where orthography is widely used. To avoid such effects, Thomson (2012) used different flags to symbolise ten different Canadian English vowels to be trained. Although work on HVPT which is explicitly related to visual modality is rare, many pronunciation applications for computers have made use of visual aids, some including ”video clips and animations of the mouth making specific sounds" (Warschauer & Healy, 1998, p. 59). Arguably, these applications may be used to train perception, although how the training in single sounds relates to listening at length is unclear.

1.2.3 Video

Video is widely used in language instruction, particularly since the advent of internet video and convenient, compact devices to play video on. While widely used, there have been few studies where the use of audiovisual and/or visual stimuli have been investigated for their potential efects on phonology acquisition.

In a laboratory-based vowel discrimination study, Masapollo et al. (2017) found that media provided in a natural audiovisual modality (i.e. where visuals are congruent with the audio) provided greater means for functionally monolingual listeners to discriminate between cross-language examples of French and English /u/ phonemes, although examples of shifts from more central to more peripheral positions were easier to perceive than shifts from more peripheral to more central. These effects were less prominent in visual-only or audio-

only modalities, thus illustrating that multimodal input can affect vowel discrimination. However, in Inceoglu (2021), L2 participants using audiovisual stimuli (a French speaker articulating and producing the target vowels) and audio-only stimuli performed approximately the same at identifying French nasal vowels, which are subject to different levels of rounding, while both groups performed better than those using visual-only stimuli. This suggests, contra Masapollo et al. (2017), that while audiovisual stimuli *can* facilitate greater discrimination, it does not always do so.

The work of Hardison (1999, 2003, 2017) on using increased visual salience for consonants in speech with gated (staged stopping) video was found to be effective with L1 Japanese and L1 Korean learners of L2 English. Her work was done with CVC words and focused on the perception of the consonants at onset and coda positions. However, in experiments carried out by Hazan et al. (2006) with video recordings, Japanese learners performed at basically chance levels in a consonant discrimination task with /l/ and /r/ regardless of visual cues, whereas Korean L1 learners of English performed at a higher level with audiovisual or audio only cues than with visual cues alone. Japanese L1 learners also performed at a chance level regardless of audiovisual or visual cues with consonants /b/ and /p/ contrasting with /v/ whereas Spanish L1 learners showed greater response to visual cues. The findings by Hazan and colleagues (2006) and by Hardison (1999, 2003, 2017) appear initially to be at odds, yet it may be the case that in gating the video, the salience of the visual cue is increased. However, for a naturally produced single word, gated video can be difficult to produce and therefore its application in classroom contexts may not be practical in comparison to simpler editing.

Similar work to Hardison's has not been carried out with regard to the visual salience of vowel quality. Hirata and Kelly (2010) investigated how mouth gestures and hand gestures affect perception of vowel length, and found that audiovisual presentations of articulation without hand gestures resulted in greater discrimination of vowel length.
“One possible explanation for this intriguing finding is that participants were “overloaded” with visual input, and this ultimately distracted them from reaping the benefits from the

mouth and lips. That is, participants had to pay attention to auditory information along with two different pieces of visual information, and this integration may have overloaded working memory such that people could not properly encode the sounds into long-term memory" (Hirata and Kelly, 2010, p. 306).

Certainly, in natural speech, hand gestures are rarely used to indicate vocalisation or other speech phenomena, but more for aspects of discourse. It therefore appears that sufficient visual information may facilitate acquisition but additional information to make speech features more salient results in less acquisition occurring.

The widespread availability of internet video media, which frequently shows articulatory gestures, has made multimodal input convenient for teachers to use. While audiovisual input may simply be convenient to source, the effectiveness of audiovisual input for phonology acquisition, particularly in internet video, has yet to be fully investigated, particularly in classroom settings. As such, the current study is highly relevant to practitioners and researchers.

An additional obstacle that may be encountered when using video in teaching and learning is misparsing phonemes due to inaccurate visual information for articulatory gestures, i.e. McGurk effects (McGurk & MacDonald, 1976). When consonants are presented in context with visual information, for listeners attending to the visual information, the signal received is the product of the acoustic and the visual information. If the visual information is inaccurate or manipulated, listeners may perceive entirely different consonants, and also holds true of L2 listeners (Iverson et al., 2003). McGurk effects can therefore contribute to L2 listening difficulties if audio and visual tracks are not correctly aligned in video, which can occur with web video (Jones, 2021), thus care must be taken to provide correctly aligned information.

1.2.4 Wave forms

A further method of exploiting visual modality is the use of wave form displays. Wave form displays have been found to facilitate phonology acquisition, (Motohashi-Saigo &

Hardison, 2009) although mainly this was investigated with regard to students specializing in linguistics. While obviously beneficial to learners whose instructors are able to produce learning materials with wave forms, an obvious disadvantage is that this cannot be used with speech in a natural or naturalistic setting because in such situations nobody actually sees accompanying wave forms.

2 Research Questions

Given that the laboratory studies reviewed above have rarely resulted in classroom-based studies that verify the ecological validity of their results, the present study addresses an existing lacuna. As such, a study among L1/dominant language Japanese learners in an intact class was conducted to investigate whether modality effects perceptual acquisition. The following research questions were investigated:

- Does visual salience affect vowel acquisition?
  - Hypothesis: visual salience improves vowel acquisition.
  - Explanation: Despite the above work by Hardison (1999, 2003, 2017), which found that increased visual salience assists the perceptual acquisition of consonants, there has been little work that is easily transferable to the classroom context which investigates the visual salience of vowels for L2+ phonology acquisition. Due to the findings in neuroscience (i.e. Berezutskaya et al., 2020; Callan et al., 2014; Glanz et al., 2018) regarding the activation of the prefrontal cortex by visual articulatory gestures and/or acoustic speech, it is expected that increasing visual salience through multimodal input facilitates greater acquisition of L2+ vowels.
- Are certain vowels affected more by visual salience than others?
  - Hypothesis: Vowels closer to the periphery of the vowel space are more easily acquired than central vowels.
  - Explanation: According to the NRV (Polka & Bohn, 2011) vowels nearer the periphery of the vowel space are more easily perceived, which can help

in 'anchoring' L2 phonological mapping. Additionally, toward the periphery of the English vowel space, there are fewer phonemes in adjacency to compete for categorisation as per the PAM(-L2) (Best, 1995; Best & Tyler, 2007).

3 Method

An intact class at a university in Japan were requested to participate in the study. All students were given an information sheet and the process of the study was explained. The study was based upon activities used for academic credit, therefore all students participated. Informed consent was requested for use of anonymised learning data, which all students provided. The institution did not require ethics board approval.

Learners were requested to take a pretest prior to being grouped and completing six listening lessons, after which a posttest and delayed posttest were administered. The grouping within an intact class means that the study methodology may be considered quasi-experimental. All tests and listening lessons were provided via a Moodle learning management system on a server hosted by the author, and taken on participants' own devices, a mixture of laptop computers, tablets and smartphones. Learners in Jones and Blume (2022) did not maintain sufficient progress with requested out-of-class work in spite of academic credit, therefore it was decided that in-class participation would be the optimal means of conducting the study. Participants were provided with lessons with the aim of improving vowel discrimination of TRAP, STRUT, THOUGHT, and NURSE.

3.1 Tests

The tests were the same as those in Jones and Blume (2022), and consisted of vowel discrimination tasks, where learners identified vowels in CVC words presented in audio-only modality, which contained either the vowel TRAP, STRUT, THOUGHT or NURSE. Vowel identification tests were chosen instead of AXB tests (i.e. tests with three items, and the middle item is the same as either the first or final presented item) due to frequent ceiling or near-ceiling values in AXB tests during pilot studies. Vowel identification tests used

graphic labels to enable easy identification of CVC words without potentially misleading learners while forgoing orthographic labels, thus maintaining test validity. Each test consisted of 32 items altogether, 8 items of each vowel, though with no more than two identical vowel discrimination items presented in adjacency. When identifying vowels, learners selected one of four photographs of simple nouns familiar to Japanese learners of English: 'cat' for TRAP, 'sun, for STRUT, 'door' for THOUGHT, and 'bird' for NURSE. The tests were identical except that the post-test and delayed post-test items were presented in a different sequence to that of the pretest. 4 weeks after the post-test was administered, a delayed post-test was administered, which was during the summer vacation, therefore it was made available for one week for participants to complete remotely away from the university campus. Test stimuli are available from the Zenodo repository (https://zenodo.org/records/11025576).

3.2 Lessons

Lessons began one week following the pretest. Participants were arranged in two groups by pairing participants with equal or roughly equal pretest scores, aiming for approximately equal mean and distribution of scores. The intervention group (AV group) was provided listening recordings in audiovisual modality, while the control group (AO group) was provided recordings in audio-only modality. Each group identified a vowel in an isolated CVC word presented by a speaker. The AV group saw the speaker's face in the whole video frame. After vowel identification, participants transcribed an utterance containing the CVC word from the prior vowel identification. Vowel identification and utterance transcription was set four times, once for each vowel to be presented. Materials from the lessons came from the Global Englishes group materials used in Jones and Blume (2022), because it was determined that there was no difference between Global Englishes and and prestige varieties for the purpose of acquisition in that study.

The recording materials were taken from TED Talks and edited for length. The AV group watched MP4 videos presented in Moodle and hosted on the author's internet server; the AO group listened to MP3 audio of the same videos, also presented in Moodle and hosted

on the author's internet server. Both video and audio were edited using Shotcut for Windows 11.

3.3 Participation

The intact class consisted of 36 students. However, four students did not complete the pretest or complete four or more lessons, therefore they were not included in data analysis. The resultant sample size after participant removal was N=32. The delayed posttest was completed by ten participants (N=10; each group n=5) during the summer vacation. All work was subject to academic credit.

4 Results

Pretest-posttest gains, and posttest-delayed posttest gains were analysed using Bayesian t-tests. Bayesian analysis was chosen because the objective of the study is exploratory rather than focused upon null-hypothesis significance testing. Bayes factors for the majority of the t-tests show no evidence for group difference in vowel learning because, according to Nourozian et al. (2019), the Bayes factors are all in the range of anecdotal strength evidence. Analysis scripts and data are available from the Zenodo repository (https://zenodo.org/records/11030325).

Table 1: t-scores and Bayes factors for gains of all vowels and individual vowels from pretest to posttest and pretest to delayed posttest.

	Posttest (N=32)		Delayed posttest (N=10)	
Description	t Score	Bayes Factor	t Score	Bayes Factor
Overall gains	-0.7040565	0.4090820	0.8774084	0.6123657
TRAP gains	0.9056559	0.4521621	0.1280369	0.4998163
STRUT gains	-2.0026242	1.4871161	0.7515674	0.6048823

	Posttest (N=32)		Delayed posttest (N=10)	
THOUGHT gains	-1.0438077	0.5027247	-0.4532984	0.5335164
NURSE gains	0.4411908	0.3673282	-1.5400906	0.9285443

Figure 1: Violin plots for gains of all vowels and individual vowels in posttest.

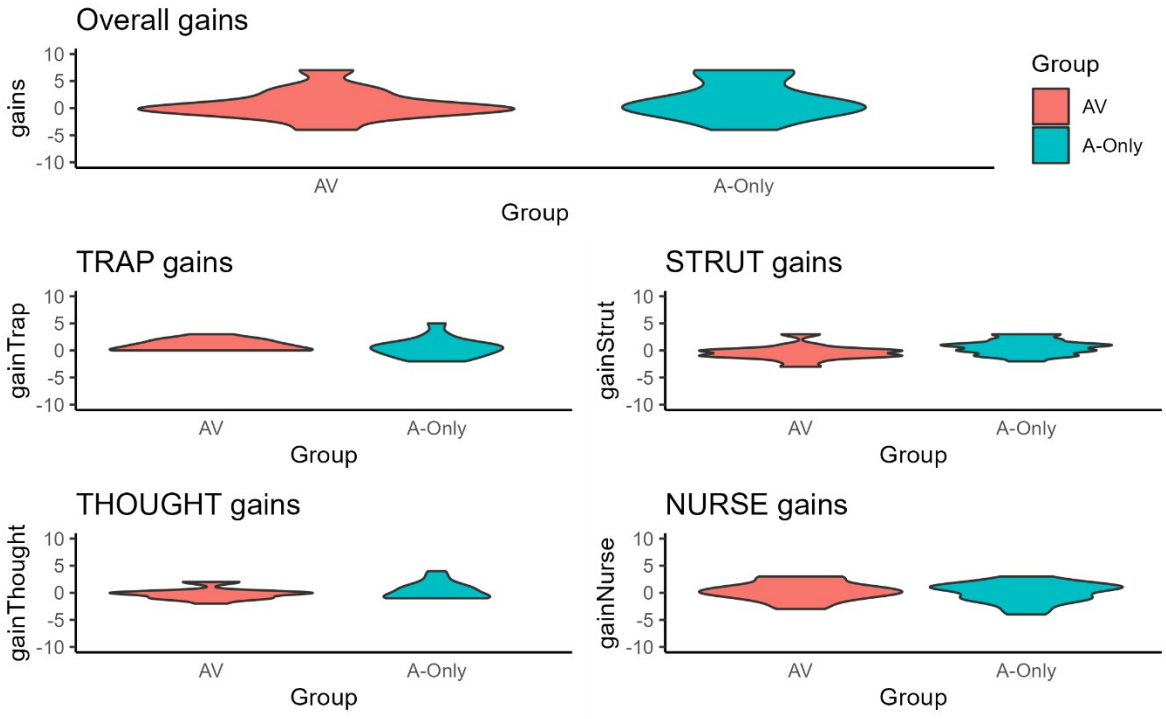
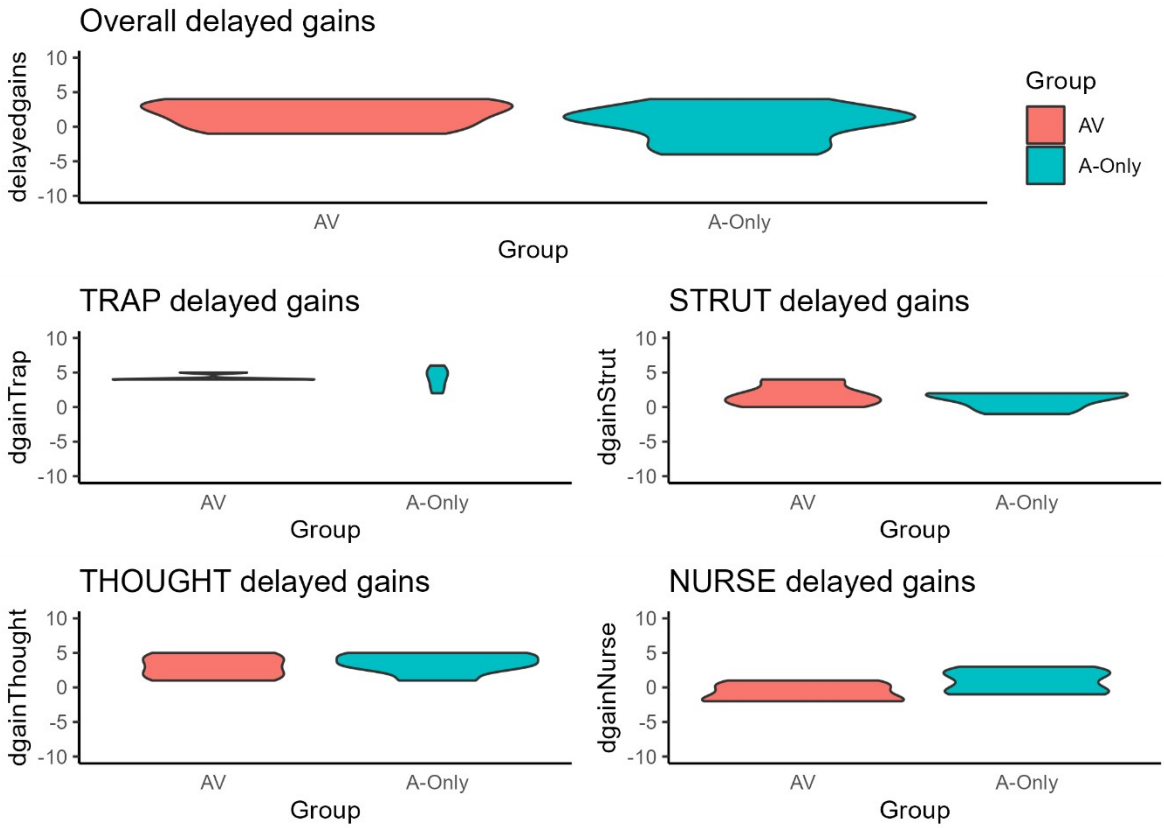


Figure 2: Violin plots for gains of all vowels and individual vowels in delayed posttest.



It is clear from the violin plots that the data is not distributed normally. However, the t-test is still appropriate because it is more powerful than non-parametric alternatives such as Wilcoxon's Signed Ranks Test and the Mann-Whitney U-Test. The violin plots also show that for the TRAP vowel gains for the AO group skew more negative than for the AV group, although there is also a cluster of higher gains for the AO group than the AV group. STRUT and THOUGHT were more difficult for the AV group to acquire than the AO group, and NURSE gains were roughly equal for both groups.

On the whole, the vowels were subject to greater gains by the AV group than the AO group in the delayed posttest. On an individual vowel basis, the AV group made greater gains on all vowels except for NURSE. The STRUT vowel was subject to the highest positive difference in gains when comparing the AV and AO groups, and TRAP was subject to more slight group gains differences. In the AO group, the NURSE vowel had higher gains

and THOUGHT slight positive gains differences, although negligible, in that fewer participants had lower end gains than the AV group.

The participants who completed the delayed post test had lower mean pretest-posttest gains and - in fact negative gains - in comparison to the participants who did not complete the delayed post test. This is also reflected in the difference for mean lesson score.

Table 2: Comparison of posttest gains: Completed delayed posttest versus no delayed posttest

Complete Delayed Posttest	Mean Posttest Gain	Mean TRAP Score	Mean STRUT Score	Mean NURSE Score	Mean THOUGHT Score	Mean Lesson Score
FALSE	1.133333	0.7666667	0.1	0.1	0.1666667	44.59207
TRUE	-1.000000	-2.0000000	1.0	1.0	-1.0000000	36.10750

Table 3: GLMM of posttest and delayed posttest gains and interaction of group condition and lesson score means and standard deviations (sds)

Gains, condition, means and sds of lesson score: interaction vs interaction	Bayes Factors
Gains, condition, means and sds of lesson score: no interaction vs interaction	1.119862e-11
Delayed gains, condition, means and sds of lesson score: no interaction vs interaction	6.898537e-02
Gains, condition, participant and individual vowel gains: no interaction vs interaction	NA
TRAP gains, condition, participant and sds of lesson score: no interaction vs interaction	1.691720e-11

Gains, condition, means and sds of lesson score: interaction vs interaction	Bayes Factors
STRUT gains, condition, participant and sds of lesson score: no interaction vs interaction	2.447049e-12
THOUGHT gains, condition, participant and sds of lesson score: no interaction vs interaction	3.144053e-12
NURSE gains, condition, participant and sds of lesson score: no interaction vs interaction	3.845319e-09
Delayed gains, condition, participant and individual vowel gains: no interaction vs interaction	Inf
TRAP delayed gains, condition, participant and sds of lesson score: no interaction vs interaction	1.645871e-12
STRUT delayed gains, condition, participant and sds of lesson score: no interaction vs interaction	3.694995e-03
THOUGHT delayed gains, condition, participant and sds of lesson score: no interaction vs interaction	6.956527e-06
NURSE delayed gains, condition, participant and sds of lesson score: no interaction vs interaction	2.231244e-03

The Bayes factors provide little evidence to support or refute interaction between gains and different factors. The two Bayes factors that stand out, for *Delayed gains, condition, means and sds of lesson score: no interaction vs interaction* and *THOUGHT delayed gains, condition, participant and sds of lesson score: no interaction vs interaction* support a theory of no interaction between these specific factors. However, all other Bayes factors should be taken as markers of merely anecdotal evidence, as per Norouzian et al. (2019).

5 Conclusion

There is nascent evidence of a difference between groups to suggest modality affects perceptual vowel acquisition but the differences do not constitute strong evidence. The participants in the AV group experienced greater gains overall in the delayed posttest, but not across all vowels. Additionally, in the initial posttest, the AV group made lower gains than the AO group. However, gains across all of the target vowels were not uniformly higher in one group when compared with another in either the posttest or delayed posttest. Furthermore, the GLMM shows, albeit weakly, that there is no interaction between delayed gains, condition, lesson score and lesson score standard deviation, thus the differences across groups must be attributed to individual differences at some level.

The lack of large difference between groups in this study shows that in classroom conditions there is possibly only a small advantage to audiovisual modality over audio-only modality for perceptual vowel acquisition. The TRAP vowel appears to be more readily learned from the audiovisual input than from audio-only input. The effect size is very small (posttest $t = 0.9056559$, delayed posttest $t = 0.1280369$) and may require reinforcement over time. The learnability of TRAP may be explained by its similarity and/or proximity to the Japanese /a/ phoneme, and its peripheral location in the vowel space, thus being more readily categorised as equivalent. In sum, for perceptual acquisition of English vowels by L1 Japanese learners, there is no immediate difference between the presentation modalities.

It may be the case that poor performance or gains motivated the participants to work harder in the period between the posttest and delayed posttest, and also to complete the delayed posttest. Unfortunately, qualitative data regarding this was not collected. However, the data here shows that prevaricating over whether audio-only or audiovisual materials are more useful for vowel acquisition is futile, as the evidence provides no robust support for audiovisual modality nor audio only, at least for relatively central vowels.

6 Discussion

The GLMM shows post-test gains are probably unrelated to lesson score mean and standard deviation. This is similar to findings by Hazan et al. (2006) regarding input modality for lateral consonant acquisition (/l-r/ contrast). While the Bayes factor is not especially large (6.90, to two decimal places), it is preliminary evidence that should be investigated further. Admittedly, this is for a very small sample size, yet it suggests that if not lesson scores, something else must facilitate the gains.

The Bayes factors show no persuasive evidence of an effect of input modality for learning gains in the posttest. Laboratory studies that provide much of the context and theory of L2 phonology acquisition largely focus on consonants. With the exception of Hirata and Kelly (2010), and Masapollo et al. (2017), the visual salience of vowels has rarely been investigated. No laboratory studies examining visual salience of L2+ vowels did so in classroom contexts, therefore whether vowel learning mechanisms in laboratory conditions transfer readily to classroom-based practice is unclear. The existing work on consonants (e.g. Hardison, 1999, 2003, 2017) points to clear benefits for acquisition from audiovisual modality. If a sufficient consonant inventory has been acquired then modality may be irrelevant to further phoneme learning, at least in the short term. However, a delayed effect for audiovisual modality may be present, although more research is required to validate any claims.

In the delayed post-test, only slightly higher gains for the AV group than the AO group were observed. The ten learners who took the delayed posttest showed motivation to take a test during their vacation time, therefore it is possible that those learners may also have undertaken independent study to remediate possible self-perceived shortcomings in their learning. Similar hypothesis is made by Mayr and Escudero (2010), who posit that reasons for learner improvements in L2 vowel perception may actually be due less to the way that learners map phonemes to categories but more related to motivation, learning context or amount of L1 and L2 use. However, regarding optimal input modality for vowel acquisition, no differences between groups to form robust evidence was observed.

Replicability of the delayed gains made by the AV group is unknown, as is the stimulus for them, hence further investigation is required.

7 Limitations

The sample size for the delayed posttest (N=10) is the clearest limitation in the current study because data analysable at a sufficiently granular level is necessary to assess factors of causality in language acquisition. While Bayesian analyses are not strictly ruled by the central limit theorem, if sample sizes are not large, the strength of evidence in the Bayes factors cannot be strong enough to suggest that effect sizes are anything beyond experimental artifacts. An additional limitation is that all learners had Japanese as L1 or as a dominant language, therefore transferability of the findings to other language pairings needs to be carefully considered. However, the ecological validity from the authentic classroom setting means that findings can inform acquisition studies in other pedagogical settings. In short, while there are certainly caveats to applying the findings, replication and extension of the current study can aid understanding of the role of multimodal input in phonology learning.

8 References

Barrios, S. L., & Hayes-Harb, R. (2020). Second language learning of phonological alternations with and without orthographic input: Evidence from the acquisition of a German-like voicing alternation. *Applied Psycholinguistics*, *41*(3), 517–545. https://doi.org/10.1017/S0142716420000077

Bassetti, B. (2023). Effects of orthography on second language phonology: Learning, awareness, perception and production. Routledge.

Berezutskaya, J., Baratin, C., Freudenburg, Z. V., & Ramsey, N. F. (2020). High-density intracranial recordings reveal a distinct site in anterior dorsal precentral cortex that tracks

perceived speech. *Human Brain Mapping*, *41*(16), 4587–4609. https://doi.org/10.1002/hbm.25144

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience.* (pp. 171–206). York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (Vol. 17, pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Bokamba, E. G. (2018). Multilingualism and theories of second language acquisition in Africa. *World Englishes*, *37*(3), 432–446. https://doi.org/10.1111/weng.12330

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*(3), 643–664. https://doi.org/10.1017/S0142716411000518

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary Size is Associated with Second-Language Vowel Perception Performance in Adult Learners. *Studies in Second Language Acquisition*, *33*(3), 433–461. https://doi.org/10.1017/S0272263111000040

Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnins.2014.00275

Escudero, P. (2005). *Lingustic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT.

Flege, J., & Bohn, O.-S. (2020). *The revised Speech Learning Model*.

Flege, J. E. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience.* (pp. 229–273). York Press.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*(01). https://doi.org/10.1017/S0272263104026117

Georgiou, G. P. (2021). Toward a new model for speech perception: The Universal Perceptual Model (UPM) of second language. *Cognitive Processing*. https://doi.org/10.1007/s10339-021-01017-6

Glanz, O., Derix, J., Kaur, R., Schulze-Bonhage, A., Auer, P., Aertsen, A., & Ball, T. (2018). Real-life speech production and perception have a shared premotor-cortical substrate. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-26801-x

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38. https://doi.org/10.1111/1467-9280.01415

Hardison, D. M. (1999). Bimodal Speech Perception by Native and Nonnative Speakers of English: Factors Influencing the McGurk Effect. *Language Learning*, *49*(s1), 213–283. https://doi.org/10.1111/0023-8333.49.s1.7

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hardison, D. (2017). Effects of Contextual and Visual Cues on Spoken Language Processing Enhancing L2 Perceptual Salience Through Focused Training. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in Second Language Acquisition*. Routledge.

Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297–326. https://doi.org/10.1017/S0261444820000658

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740–1751. https://doi.org/10.1121/1.2166611

Hirata, Y., & Kelly, S. D. (2010). Effects of Lips and Hands on Auditory Learning of Second-Language Speech Sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310. https://doi.org/10.1044/1092-4388(2009/08-0243)

Inceoglu, S. (2021). Language Experience and Subjective Word Familiarity on the Multimodal Perception of Non-Native Speakers' Vowels. *Language & Speech*, 1. https://doi.org/10.1177/0023830921998723

International Phonetic Association (Ed.). (1999). Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Jones, M. (2021). Problems Teaching Listening Online. *The Language Scholar*, *9*, 74–83. https://languagescholar.leeds.ac.uk/problems-teaching-listening-online/

Jones, M., & Blume, C. (2022). Accent Difference Makes No Difference to Phoneme Acquisition. *Teaching English as a Second or Foreign Language Journal--TESL-EJ*, *26*(3), 1–22. https://doi.org/10.55593/ej.26103a3

Lenneberg, E. H. (1967). Biological foundations of language. Wiley.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279–297. https://doi.org/10.1017/S1366728909990022

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), Article 5588. https://doi.org/10.1038/264746a0

Mora, J. C., & Fullana, N. (2007). Production and perception of English /i/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. *ICPhS XVI*, 1611–1616. https://www.researchgate.net/profile/Joan_C_Mora/publication/242527090_Production_and_perception_of_English_i-I_and_ae-VV_in_a_formal_setting_Investigating_the_effects_of_experience_and_starting_age/links/551bc4db0cf2909047b961f0/Production-and-perception-of-English-i-I-and-ae-VV-in-a-formal-setting-Investigating-the-effects-of-experience-and-starting-age.pdf

Mora, J. C., & Mora-Plaza, I. (2023). From Research in the Lab to Pedagogical Practices in the EFL Classroom: The Case of Task-Based Pronunciation Teaching. *Education Sciences*, *13*(10), Article 10. https://doi.org/10.3390/educsci13101042

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*. https://doi.org/10.1515/phon-2022-2018

Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, *13*(2), 29–47.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496–1509. https://doi.org/10.1044/1092-4388(2007/103)

Nishi, K., & Kewley-Port, D. (2008). Non-native Speech Perception Training Using Vowel Subsets: Effects of Vowels in Sets and Order of Training. *Journal of Speech, Language, and Hearing Research : JSLHR*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Norouzian, R., de Miranda, M., & Plonsky, L. (2019). A Bayesian Approach to Measuring Evidence in L2 Research: An Empirical Investigation. *The Modern Language Journal*, *103*(1), 248–261. https://doi.org/10.1111/modl.12543

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*(1), 201–225. https://doi.org/10.1017/S0272263117000432

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Shirai, Y. (1990). U-shaped behavior in L2 acquisition. In H. Burmeister & L. P. Rounds (Eds.), Variability in second language acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum, Vol. 2 (pp. 685–700).

Sueyoshi, A., & Hardison, D. M. (2005). The Role of Gestures and Facial Cues in Second Language Listening Comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Tyler, M. D. (2019). PAM-L2 and Phonological Category Acquisition in the Foreign Language Classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 607–630). Aarhus University.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, *31*, 57–71.

Appendix 4

Jones, M. (2024d). *Listening to Global Englishes: Problems in practice* [Manuscript submitted for publication].

Listening to Global Englishes: Problems in practice

Marc Jones

Attention is currently focused upon global varieties of English as more English language learners are being exposed to a wider range of varieties. This study reports findings from qualitative data regarding listening difficulties taken from university students in Japan, whose first/dominant language is Japanese. Of 16 learners, 8 listened to prestige variety speakers and 8 listened to Global Englishes (i.e. varieties of English spoken in post-colonial communities or communities in which English is not a main language), in edited TED Talks. No meaningful difference was found between groups regarding reported ease or difficulty with speakers' English. Participants reported that unfamiliarity with varieties of English caused comprehension problems, including American English, a highly graded version of which is the main variety taught in the Japanese school system. Additionally, content and vocabulary also proved challenging for participants.

只今、学習者が広い英語種類の範囲に接されているので国際的な英語の種類が注目されている。この論文が聴解困難について定性的なデータを述べる。データの元は第一言語の日本語使い日本にいる大学生である。16 人が編集せれた TED Talks を聞いた。16 人の中で、8 人が名声種類英語、8 人が国際種類英語を聞いた。簡単さや難しさが特別の差がなかった。見慣れない種類で聴解困難が起こったといっても、小中高校が教えられたアメリカ英語も聴解困難があった。それ以上、内容も語彙も学習者には難しかった。

Global Englishes for listening

Given that many L2+ English learners are likely to interact with a range of interlocutors speaking different varieties of English. Although learners may consider that future interactions will be with speakers of non-standardised varieties, they may not know what features of those varieties to prioritise in their learning. The obstacles preventing instruction of Englishes beyond the status quo of prestige varieties are not insurmountable. GE varieties are present in the wealth of authentic materials available on the internet. It can be easy to see which varieties are available, through speaker biographies (e.g.,TED Talks), or flag icons (e.g. Elllo.org). Learners as well as teachers have agency to make decisions about learning materials, and they can "develop their GE knowledge, harness their GE self-efficacy, develop their English skills, and enhance their critical awareness of GE" by doing so (Widodo et al., 2022, 195), thus GE listening is highly practical. However, while learners have access to GEs, they may encounter problems in initial comprehension and therefore develop apprehension toward listening to varieties of English that are less prevalent in mainstream English materials. Additionally, some teachers may also lack experience in teaching with GEs and could be unsure what to prioritise for instruction. If it is possible to consider what learners may find difficult in listening to GEs and prestige varieties (PVs), comparisons can be made and more efficient learning and teaching is possible.

Global Englishes

English standardised PVs are assumed to be useful as classroom models. Baratta (2019) advocates the use of a standardised PV as a classroom model with awareness of different GEs raised through exposure. Such an approach risks relegating GEs as an afterthought, and may even maintain the hegemony of PVs over GEs. To avoid tokenisation or exoticisation, Widodo et al. (2022) recommend that materials used to teach GEs are authentic and from a range of cultural contexts. Additionally, when teaching GEs, it is

important not only to contrast GEs only with standardised varieties, but also with other GEs and instances of English used where speakers of differing varieties accommodate one another, and not to conflate GEs with English used internationally by speakers of differing L1s (Honna, 2008). Teachers can then instruct learners in how to best orient to speakers in order to decode speech more effectively.

A further issue with standardised PV models is that of othering non-white speakers of English. Globalised ELT materials maintain a status quo in which the prevalent varieties found are those of ostensibly white speakers of the UK and US (Rubdy, 2015). This situation perpetuates '(Non)Native Speakering' (Aneja, 2016), a process of othering due to racial characteristics and some speech differences, whether English is the subject's main/dominant language or not. (Non)Native Speakering discourse is related to prescriptivist language norms, and occurs "in terms of explicit speech acts (e.g., 'You are not a native speaker') or implicit discursive positioning (e.g., in asking a person of color how they learned English so well)" (Aneja, 2016, 364). The latter part of Aneja's description illustrates clearly that racial factors rather than language proficiency is the issue. A matched guise study, by Hartmann (2021) found that PV speakers are rated more positively than racialised speakers. Without classroom work using GEs and racialised speakers such detrimental biases will arguably be maintained.

In initial work to investigate the effects of using GEs for phonology acquisition, as opposed to the prestige-variety status quo, Jones and Blume (2022) in a small-scale study found no effect on vowel learning when comparing the effects across groups using self-access listening materials. However, given the way that speakers of varieties that are afforded less prestige are marginalised in ELT materials (Rubdy, 2015), encouraging greater access to a wider range of Englishes is more likely to facilitate the normalisation of GEs.

One of the difficulties faced by listeners is that parsing speech involves dealing with the great variability in phonemic articulation of segments, even by a single speaker (Cutler, 2012). That is, a given speech sound, for example /s/, can differ in quality depending on many factors involved in producing contemporaneous speech. When considering a speech

community, the variation increases further. When speech communities become superdiverse, as is arguably the case with English(es), the L1+ background of speakers can also influence phonology (Gut et al., 2023).

Prestige Varieties

Greater familiarity with a given variety should enable perceptual acquisition and render comprehension easier. However, structures that learners are unaccustomed to hearing may take longer to acquire than those similar to L1 (Archibald, 2021), as is also the case for discrimination of more complex phonological inventories (for example several L2 vowels in proximity to many fewer L1 vowels; Darcy et al., 2012). Archibald (2021) suggests that "structures which can be parsed are easier to acquire than structures which the parser cannot yet handle" (p. 3), or in lay terms, *if you can't hear it, you can't get it,* although this can change with exposure over time. Therefore, if the target phonology is not developed sufficiently to parse phonemes effectively, listening is hindered. Such difficulties are attested to in Galloway (2013), where student participants reported unfamiliarity and difficulty understanding British English in spite of being third and fourth-year English majors at university, and also taught by Galloway herself, a British English speaker.

Furthermore, learner identifications of accent can be highly inaccurate. In a study conducted in an English for Academic Purposes unit at a university in Scotland (Archer, 2022), learners identified Standard Southern British English as General American, and also displayed "an apparent lack of awareness of the geographical and phonological differences which exist in the individual nations within the UK" (p. 76). What this means for learners listening to PVs is that not only do learners potentially misidentify PVs, but it may also be the case among learners who indicate a preference for PVs due to an ease of listening that they are basing their judgement on preconceptions, racialisation and social prestige rather than acoustic or psycholinguistic factors.

However, just as there is no such singular thing as Standard Southern British English, nor is there such a thing as one 'Global English'. Instead, a deeper understanding of how language works is required. This understanding is not particularly onerous and may actually lead to better outcomes in language learning, because there will also be a deeper understanding of what is real and what is, in some EFL materials and teaching contexts, perhaps merely a convenient fiction that allows syllabus coverage with no questions asked. By considering language as a set of practices that can be understood by means of exposure and instruction, and eventually acquired, there is a higher likelihood of observing acquisition and learners maintaining long-term motivation.

The Study

Aims

In the current study, qualitative self-report data regarding the relative difficulty of listening to different English varieties in TED Talks as video texts is examined.

Are speakers of GEs perceived as more difficult to comprehend than PV speakers?

Racialisation in listening is attested to by Hartmann (2021). That is, white (or white-passing) is normalised. While difficulty in listening is sometimes actually difficulty in listening, whether there will be differences between groups watching GEs or PVs TED Talks is thus far unattested. The benefits of investigating the difficulties between GEs and PVs are that teachers and learners can be made aware in advance of difficulties across and between certain variety types and can consider how best to approach the listening process depending upon the variety spoken.

What aspects of the videos do the learners report as difficult, if any?

Learners usually attest to speed of speech as being difficult to parse (Renandya & Farrell, 2010), but difficulties may actually result from connected speech, which may not be directly instructed (Siegel, 2016). Unknown content vocabulary was expected to be cited as a difficulty for both GEs and PVs. However, the difficulties reported by the groups were unknown prior to data collection. Understanding what learners find difficult regarding speakers, content and variety can aid understanding of learner needs in listening generally, and also for English for academic purposes in particular. However, it is unknown whether the difficulties will be reported for TED Talks, because they are well known as English materials suitable for learners in Japan, and which many learners use for independent study, albeit often with subtitles.

Institutional context

The current study was carried out at the author's institution, a private university in Tokyo. Within the institution, the author teaches English for academic purposes courses in order to prepare first-year students for their English-medium instruction courses. The sample of participants was chosen because they were part of an intact class of 23 that was taught by the author. The tests and lessons were set as independent out-of-class study for academic credit, and 16 students completed the entire set of tests and lessons and returned completed consent forms. All participants had at least six years of prior English study, with some participants also having prior foreign travel and study abroad experience. Participants were grouped in two conditions: one group watched TED Talks with GEs speakers (GE) and one group PV speakers (PV). Informed consent was obtained and participants were informed that their anonymised data would be published. Participants were paired by their scores in a pretest: participants attaining the same score or similar placed in separate groups to ensure

that group pretest scores were approximately equal. The qualitative data for learners completing the entire self-access course of five weeks plus post-test consisted of a final sample of 16, with eight in each group.

Study Design

Data was collected from participants while collecting quantitative data for Jones and Blume (2022). Data was obtained through a self-paced online course through Moodle, which supplemented a regular university course. Qualitative data was elicited after participants had completed a) four rounds of vowel identification followed by an utterance transcription, and b) a summary of the whole videotext. Participants were asked to comment on the difficulties that they encountered in listening. The staging of data collection means that the participants could adequately reflect on the challenge of completing the vowel identifications, utterance transcriptions and the summary. While rich data was anticipated due to the participants' rapport with the researcher, who was also their classroom teacher, the Moodle lesson was structured to computationally permit single word or even single character answers to the data collection question for each video. However, all of the participants provided detailed answers for each video they viewed. The qualitative data obtained regarding perceived difficulty was analysed for count data and collocations to assist in analysis of the data through post-hoc interpretive coding.

The study was conducted online due to the COVID-19 pandemic, during which sporadic studying and working from home was necessary. This partial online-mediated communication could have impacted the process of the research and hence results in different ways. By providing the materials online, the participants had relatively easy access and difficulties caused by playback 'glitches' transmitting high-quality prerecorded video over videoconferencing software can be minimised. Technical problems affecting listening were not reported by participants.

Moodle was chosen to mediate the delivery of the videos for several reasons. The types of learning management system (LMS) available (Manaba and Google Classroom) were inadequate for the tasks. While videos can be accessed in both Manaba and Google Classroom they require delivery through either YouTube or Google Drive, neither of which can facilitate the delivery of specific excerpts. Furthermore, when using YouTube for delivery, advertisements may be served which is unsatisfying scientifically, aesthetically and ethically.

Results and Discussion

Learner perceptions of GEs

An unsurprising finding is that the learners found listening to a speaker of Indian English difficult, for example:

The difficulties were understanding pronunciation with Indian accent. I felt challenge to understand what she said the word because I am not used to the Indian sound. I need practice to clearly understand the Indian-English. (Participant 8, GE).

Note that Participant 8 makes explicit that they attribute their difficulties not to speaker-derived factors but to a lack of prior exposure to Indian English. Interestingly, they also identify a need for further exposure, perhaps due to the university department's focus on globalisation and global issues.

Two participants in the GE group made comments about the content of texts. Relating to the Chinese speaker, the “topic was interesting which made me want to listen him carefully” (Participant 2 GE); and relating to the Kenyan English speaker, “I totally

concentrated on this video since the content is quite interesting without feeling it is boring." (Participant 4 GE).

These comments relate to the orientation toward the speakers and the texts due to content, thus fostering motivation. As Participant 4 (GE) states, "I want to be able to figure out at least 90 percent without another sentence." Such motivation is clearly advantageous for teaching and learning, leading to more listening and arguably likely to result in more acquisition.

Learner perceptions of PVs

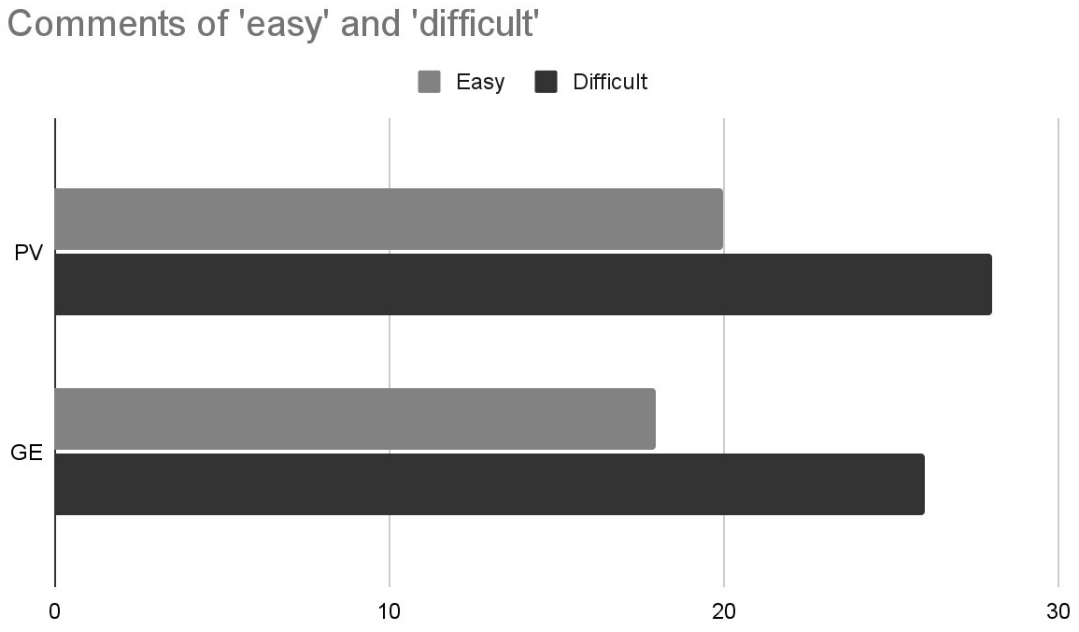
Two participants from the PV group attributing difficulties to British English speakers. For example "Her English has pretty strong British accent, which sometimes make me to think carefully so it was quite useful" (Participant 5 PV). This attribution of difficulty to British English is interesting, particularly for a group of L1 Japanese participants because it echoes comments by Galloway (2013) that her students/participants also found British English difficult. However, the difficulty here is commented by Participant 5 as being useful.

American English is a variety that is supposedly prevalent in school English materials in Japan. In spite of this, the white speaker Eastern American English speaker, arguably a hegemonic model, gained the greatest number of comments with the word 'difficult' from the PV group. Participant 15 (PV) makes this particularly clear: "In terms of video, the content was not so difficult but the way he speaks was unusual, therefore it was arduous for me." Arguably, most educated Americans from the East Coast would assume that this variety of English to be highly comprehensible by English users worldwide, which is not the case.

Similarities and Differences

In participants' comments regarding the videos in the lessons, there were 34 instances of the words 'difficult' or 'difficulty' by the PV group compared to 27 comments by the GE group. When these numbers are adjusted for comments including 'not difficult' or about 'difficult situations' faced by speakers in the videos, 28 PV group members refer to difficulties in comparison to 26 by the GE group. When observing the use of the word 'easy', the PV group had 20 incidences compared to 18 incidences in the GE group. Both the differences between the PV and GE groups for 'difficult'/'difficulty' and 'easy' are negligible. Only one participant in the PV group contradicted themselves regarding ease and difficulty.

Figure 1: A graph of the PV/GE group frequency of easy and difficulty relating to listening texts. Each group: n=8, text set: n=5.



Speed of speech was also mentioned across both groups, with five mentions of the word ‘slowly’ from the GE group; there were six mentions of the word ‘fast’ from the PV group and three mentions of ‘fast’ from the GE group. The clear observation here is that there is a difference between the groups in mentioning speed of speech. Admittedly, perceived speed can be related to factors other than number of syllables or words per minute (Henrichsen, 1984). However, in the participants' answers the word ‘slowly’ is related to their listening success, while the word ‘fast’ is related to their listening difficulty. The implications for this are discussed in the section below.

Conclusion

Summary of findings

There is no meaningful difference between groups for statements of ease or difficulty in comprehending the varieties, therefore stating a case for PVs being easier than GEs or vice versa is impossible. Participants’ comments regarding vocabulary and content difficulty leads to a conclusion that the difficulties faced in listening to GEs and PVs of English are fairly universal. The common denominator is a lack of exposure to different English varieties to facilitate comfortable comprehension.

Aspects of the videos that learners reported as difficult included accent, speech rate, content and vocabulary. Accent was mentioned as a factor contributing to difficulty or ease by some participants but was equal across groups. Thus, providing greater familiarity to varieties that are most likely to be encountered by learners is paramount. Even British and American varieties common in English-language mass media caused difficulties for the participants, therefore other varieties may require even more repeated exposure. Speech rate was mentioned by learners in relation to both difficulty and ease, although high speed

causing difficulties was mentioned only in connection with PV speakers. As expected, content and vocabulary were mentioned by participants in both groups, because the video topics were diverse in order to provide suitable input for university students in an interdisciplinary degree programme. However, this rarely proved very problematic for either group.

Teaching implications

In spite of the supposed familiarity that learners in Japan have with PVs of English, particularly American English, PVs were no easier to listen to than GEs, which are arguably less familiar to learners, suggesting that language in learning materials bears little resemblance to the natural rehearsed speech used in TED Talks. Graded language in materials may result in a lack of input available for natural language, and also be demotivating: overly simple language use may foster a false sense of achievement and ability, therefore when authentic language is encountered, learners' abilities are found wanting, thus rendering prior learning to be perceived as pointless. Interestingly, GEs may be more learner friendly than graded speech based upon PVs, because they can orient listeners to content and also foster a greater disposition toward intercultural communication.

These findings support a reasonable opinion that unfortunately needs to be repeated: the absence of GE speakers from ELT materials is a travesty that has been excused time and again as being in learners' interests. Excuses about relative ease of different varieties are faulty heuristics biased toward notions of social prestige and cultural hegemony. Varieties to be instructed need to be chosen on the communication needs of learners and phonological differences need to be taught. Nobody is arguing to stop teaching PVs, only to ensure that GEs are also taught in a non-tokenised way.

Limitations

The study was conducted with first-year university students in an EMI programme with a global orientation, and therefore the participants may have been more amenable to listening to GEs than other populations. It should be pointed out that the PV group also received GEs input after data collection for the study ended. A further limitation to the study is that the use of TED Talks provides learners with erudite, well-rehearsed speakers, whichever variety of English is spoken. Documentaries and realist dramas may provide more difficult input, even if more lexical items are known prior to viewing. Further research in this area is recommended. Finally, a quantitative study of listening difficulty was not undertaken with the learners. Arguably the data received from the learners in the current study was richer and more meaningful than self-reported ordinal-scaled questionnaire measures with such a small sample.

References

Aneja, G. A. (2016). Rethinking Nativeness: Toward a Dynamic Paradigm of (Non)Native Speakering. *Critical Inquiry in Language Studies*, *13*(4), 351–379. https://doi.org/10.1080/15427587.2016.1185373

Archer, G. (2022). The effects of prestige model familiarity on students' perceptions of and interactions with diverse English accents. In Sardegna, V. G. & Jarosz, A. (Eds.), *Theoretical and practical developments in English speech assessment, research and*

training: Studies in honour of Ewa Waniek-Klimczak (pp. 67–85). Springer. https://doi.org/10.1007/978-3-030-98218-8_5

Archibald, J. (2021). Ease and difficulty in L2 phonology: A mini-review. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.626529

Baratta, A. (2019). *World Englishes in English language teaching*. Springer. https://doi.org/10.1007/978-3-030-13286-6

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English– L2 French acquisition. *Second Language Research*, *28*(1), 5–40. https://doi.org/10.1177/0267658311423455

Galloway, N. (2013). Global Englishes and English Language Teaching (ELT) – Bridging the gap between theory and practice in a Japanese context. *System*, *41*(3), 786–803. https://doi.org/10.1016/j.system.2013.07.019

Gut, U., Kopečková, R., & Nelson, C. (2023). *Phonetics and phonology in multilingual language development*. Cambridge University Press. http://dx.doi.org/10.1017/9781108992527

Hartmann, J. (2021). Tomorrow's teachers' perceptions of Global Englishes. In Callies, M., Hehner, S., & Meer, P. (Eds.), *Glocalising teaching English as an international language: New perspectives for teaching and teacher education in Germany* (pp. 46–62). Routledge. https://doi.org/10.4324/9781003090106

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, *34*(3), 103–123. https://doi.org/10.1111/j.1467-1770.1984.tb00343.x

Honna, N. (2008). *English as a multicultural language in Asian contexts: Issues and ideas* (1st ed). Kuroshio Publishers.

Jones, M., & Blume, C. (2022). Accent difference makes no difference to phoneme acquisition. *Teaching English as a Second or Foreign Language Journal--TESL-EJ*, *26*(3), 1–22. https://doi.org/10.55593/ej.26103a3

Renandya, W. A., & Farrell, T. S. C. (2010). 'Teacher, the tape is too fast!' Extensive listening in ELT. *ELT Journal*, *65*(1), 52–59. doi:10.1093/elt/ccq015

Rubdy, R. (2015). Unequal Englishes: The Native Speaker, and Decolonization in TESOL. In R. Tupas (Ed.), *Unequal Englishes: The politics of Englishes today* (pp. 42–58). Palgrave Macmillan. https://doi.org/10.1057/9781137461223_3

Siegel, J. (2014). Exploring L2 listening instruction: Examinations of practice. *ELT Journal*, *68*(1), 22–30. https://doi.org/10.1093/elt/cct058

Widodo, H. P., Fang, F., & Elyas, T. (2022). Designing English language materials from the perspective of Global Englishes. *Asian Englishes*, *24*(2), 186–198. https://doi.org/10.1080/13488678.2022.2062540

Appendix 5

Jones, M. (2024e). *Feedback timing affects L2+ perceptual vowel acquisition.* [Manuscript submitted for publication].

Feedback timing affects L2+ perceptual vowel acquisition

Marc Jones

Toyo University, Japan; TU Dortmund, Germany

ABSTRACT

L2+ English vowel perception can be difficult for listeners whose first languages do not have as wide a vowel inventory. This study examines the role of feedback in acquisition of English vowels for perception. The current study was conducted within an intact class at a university in Japan, with a total 34 participants completing sufficient work for inclusion in the study data. After a pretest, two groups took vowel discrimination lessons over four weeks, followed by a post-test one week after those lessons concluded. One group received immediate verbal feedback while another received delayed verbal feedback. Data were analysed through Bayesian t-tests and generalized linear mixed models (GLMM) in order to explore the major factors in pretest to post-test gains. The t-tests resulted in low scores for all vowels, though delayed feedback was favoured. In the GLMMs, group factors contributed more movement to models than timing effects on lesson completion intervals. While the results are somewhat inconclusive regarding which feedback modality is most affective for acquisition the ecological validity of the study being conducted in a classroom means that the transferability of findings is possible, and that the methodology can be repeated across contexts.

KEYWORDS: English, feedback, listening, phonology acquisition, vowel

INTRODUCTION

In the English Language Teaching literature, Jenkins (2000) states that vowel quality is not important for speakers to be understood; however, in order to ensure that language learners reach their potential, sufficient instruction of listening is necessary in order to bring about a more complete phonology which should make the listening process easier. One of the key was that language tends to be taught is through feedback.

This article reports on the findings of a study investigating the effect of feedback timing on perceptual vowel acquisition, that is, the ability to discriminate between vowels. The study was conducted with an intact class at a university in Tokyo. Learners were tested and taught the vowels /æ/, /ʌ/, /ɜ:/, and /ɔ:/ using a Moodle installation run by the author. Data was collected in face-to-face lessons in 2023 and then analysed between 2023 and 2024. Effects are reported based upon Bayesian analyses of pretest and post-test scores as well as class attendance and/or making up of work, the importance of which will become clear as participant attrition is discussed.

LITERATURE REVIEW

Vowel learning

Learning to understand L2+ speech is important for hearing language learners. The goal for most learners is generally communication, yet if listening skills are suboptimal then communication is hindered. If the target language phonology remains incomplete and is not addressed, the result is that learners may become demotivated due to a perceived lack of competence (Ryan & Deci, 2017). Demotivation may then lead to a cycle of lowered attention to instruction and thus failure to learn; participation in classes where learning does not occur may even result in foreign language anxiety (Horwitz, 1986, yet see also Rebuck, 2008 for an opposing perspective). Obviously such conditions should be avoided as much as possible, which makes the case for providing phonology instruction and/or improving the conditions for phonology to be acquired.

The Perceptual Assimilation Model (PAM) (Best, 1995) was developed in order to explain how a naive listener to an unfamiliar language may categorise the speech sounds in relation to L1: same as L1 category, good fit to L1 category, poor fit to L1 category, uncategorized according to L1. PAM-L2 (Best & Tyler, 2007; Tyler 2019) built further upon the model, yet posits that if L1 categorizations allow for successful discrimination of L2 phonemes, learning an L2 categorization is unnecessary because perception occurs at the level of gesture and phonetic level, up to a general phonological abstraction. On the other hand, if L1 categorization does not allow learners to effectively discriminate between L2 phonemes, perceptual learning is necessary and therefore categorization needs to take place. Such categorization takes place on the same template as L1, i.e., the phonological template is common to all languages. It is assumed that without a large

amount of input, estimated at a term of 6 months in an environment where the language is widely spoken (Best & Tyler, 2007), or much longer in a foreign language learning environment, that L2 learners do not easily develop a phonological model that would enable accurate perception or pronunciation of novel L2 categories. However, Tyler (2019) clarifies that "it is not necessarily a lack of native-speaker input that may reduce the likelihood of L2 category acquisition in the classroom, but input that fails to provide clear phonological differences between L2 categories." (p.616)

However, not all phonology scholars are in accord regarding the PAM-L2. For instance, the Universal Perceptual Model (UPM) (Georgiou, 2021) states that learners do not have phonological categories but a "phonological space [which] is filled with units called phonetic categories" (p. 4), and Mitterer, Reinisch and McQueen (2018) propose that learners internalise context-specific allophones as opposed to context-independent phonemes. Regarding the PAM-L2 agnosticism regarding the link between perception and production, Gut, Kopečková and Nelson (2023) state that their results of a study involving L3+ learning of liquids suggest that perception and production *are* linked.

The Speech Learning Model (SLM) (Flege, 1995) and Speech Learning Model revised (SLM-r) (Flege & Bohn, 2020) state that learners in a language environment long-term, i.e. migrants, can develop phonetic categories associated with their L2 independent of L1, and that over time these categories may blur together when they are in proximity (Flege, 1995). However, while the initial SLM hypothesised that age of onset of learning predicts a difficulty in forming separate L2 categories for similar L2 sounds that would be categorised the same in L1 (Flege, 1995), the SLM-r states that there is "evidence that Late learners can gain access to L1-L2 phonetic differences, store the detected differences in long-term memory, and then use the stored perceptual representations to guide articulation" (p. 8). With this being the case, adult learners should be able to develop speech categories, but perhaps using different pathways to early learners due to experience and existing learned behaviour.

However, the PAM and SLM both have limits to their application. Faris et al (2018) investigated the acquisition of L2 Danish perceptual vowel categorisations by naive users, who were L1 speakers of Australian English. Where Danish vowels overlap with Australian English vowels, it was found that

participants generally categorised the Danish vowels upon input with one of the L1 Australian English categories. The authors state that their findings suggest "the use of an arbitrary assimilation criterion may not sufficiently reflect systematic responses that are below the categorisation threshold" (Faris et al., 2018, p. 9). That is, predicting categorisations based on formant frequencies may not always be possible. Faris and colleagues' findings are similar to those of Ingram and Park (1997), who investigated L1 Japanese and Korean listeners of Australian English, and found that Japanese listeners categorised Australian vowels using L1 categories with speaker-dependent duration. As such, teachers and learners need to be aware of the difficulties when studying languages with a larger vowel inventory than their L1 and deliberately pay attention to difficulties in parsing speech, in order to make eventual moves toward categorisation and discrimination of phonemes. A further consideration to be made is that, according to Polka and Bohn's (2011) Natural Referent Vowel framework, L2 vowels that are in a more peripheral position in the vowel space (i.e. very closed or very open, and with very back or very front tongue) are more likely to be learned, because they are closer to L1 vowels that are basically universals which act as referents for other vowels in the vowel space.

The Second Language Linguistic Perception (L2LP) model (Escudero, 2005; Van Leussen & Escudero, 2015) posits that when adult listeners of an L2+ perceive speech sounds they perceive them at a pre-lexical level. This is the same as in Optimality Theory (OT; Prince & Smolensky, 2002; see Archangeli, 1999 for an overview); that is, the sounds are not immediately mapped to lexical candidates, but instead are mapped to a duplicate L1 phonological template (Van Leussen & Escudero, 2015), that is, phonological categories are revised as the L2 is learned. Escudero (2005) claims "a separation of perceptual mappings from sound representations leads to an adequate comparison of the perception systems involved" (p. 86). As such, the model diverges from the PAM, which models the phonological template as being shared between L1 and L2+, with different, changing L2 categorisations on the same template over time as the language is acquired, rather than the L2LP and OT's duplicate.

The key states of L2 in the model are the L2 initial state, L2 development, and the L2 end state. While these key states are logical, they may be problematic in their operation. L2 initial state is essentially whatever state a learner is in at the start of learning or when they are of concern to a

researcher. However, the L2 development and L2 end states may not be so easily discerned. At the end of an L2 acquisition study, is the learner-participant in an end state or a development state? How is the end state identified? Can the L2 development state be wrongly identified as being the L2 end state, due to an insufficient amount of exposure and/or use of language in order to move to a higher state of development or the end state (c.f. Rastelli's 2016, theory of quantum input)? Arguments continue about what exactly is happening in language acquisition, and even in phonology acquisition. Yet even while research mounts, there are still uncertainties in causal mechanisms and catalysts, thus more research is necessary.

Therefore, in the L2LP, perceptual representations of speech sounds are claimed to result from constraints of the perceptual mappings at the initial state. As exposure is increased to the language, the L2 development state occurs until an end state is reached. In simplified terms, mental categories need the speech signal to actually follow the constraints, else perception is inaccurate and the speech sounds are not categorised and therefore not parsed (Escudero, 2005). However, with deviations from the constraints, perhaps due to an unknown variety or dialect, perceptual representations may be somewhat malleable, as found in Williams and Escudero (2014).

According to Flege and MacKay (2004), there can be difficulties in learning vowel contrasts for according to age of onset. However, the outcomes of L2 phonology acquisition are not straightforwardly linked with age: "[B]eginning to learn an L2 in childhood does not guarantee a nativelike perception of L2 vowels. On the other hand, establishment of the L1 phonetic system does not guarantee that measurable vowel-perception differences will exist between early learners and L2 native speakers" (p. 26). Obviously, individual differences factor into the acquisition of perceptual phonology acquisition. Differences that Flege and MacKay attribute to acquisition of perception are years of education in the L2 community overall (as opposed to age of arrival in the L2 community), and percentage of L1 spoken. Baker et al. (2002) also attributed age as a factor because "child L2 learners are less likely than adult L2 learners to judge L2 sounds as members of L1 sound categories" (p. 45), although in the SLM-r (Flege & Bohn, 2020), note that age itself need not be a limiting factor in learning an additional language, but is likely to interact or be confounded with other variables.

One of the important aspects of listening is that, through perception, and in turn attention, the acquisition of language can be facilitated. In doing so, there is an interaction between the act of listening, and acquisition of linguistic items. The Vocab model (Bundgaard-Nielsen et al., 2011, 2012) states that lexical acquisition interacts with phonological acquisition. The categorisation of phonemes allows for the accurate decoding of syllables, and lexis. Additionally Felker and colleagues (2021) show that lexical knowledge mediates learning of vowel categorization for perception of an interlocutor's speech. However, Escudero's (2005) LP model posits that "perception is is pre-lexical because it is the result of the pre-lexical decoding of the speech signal" (p. 51) with codes used for lexical access after perception.

Escudero's arguments for perception being a precursor to lexical decoding appear to be sound, based upon prior studies (Baese-Berk, 2019; Crain et al., 2017). Despite the differences betweeen L2LP, PAM-L2 and SLM-r, it is possible to consider an intermediate stage between perception and lexical decoding where perceived phonemes are given preliminary assignment to lexis, pending verification. Such preliminary assignation may occur as a result of language learners/users testing their personal hypotheses (as with lexicogrammar in Processability Theory, Pienemann, 1999) and experiencing either a perceived successful listening event (through perceived successful comprehension, which may or may not be verified by an interlocutor), or an unsuccessful listening event. In both cases, the perceived phonemes (either correctly or incorrectly perceived) are only one factor in the (lack of) listening success, the others being lexical knowledge, syntax knowledge, (Vafaee and Suzuki, 2020) and access to semantic knowledge (Rukthong & Brunfaut, 2020). As such, perceptual phonology may assist in the lexical acquisition process through listening. It is thus the purpose of the study to consider perceptual phonology acquisition and assume that every learner has some differences in their L2+ lexicon.

Feedback

When considering corrective feedback for SLA, many of the findings are described in ways that appear to be universal across syntax, lexis, and phonology systems, and yet this may not be true. Listening and perceptual phonology in particular may require deeper consideration in the feedback to be implemented due to the fact that output may not be a reliable proxy measure of

phonological perception (Winn & Teece, 2021). Thus, greater consideration of assessment procedures are required.

When considering instructed SLA, Lyster and Ranta (2013) stated that “Recasts and explicit correction can lead only to repetition of correct forms by students, whereas prompts can lead either to self-repair or peer repair, not to repetition” (p. 172). However, this may not always be true regarding phonology instruction. Although learners may not always perceive recasted or explicit phonological feedback correctly it can provide more input, whereas a prompt may allow learners to realise that their phonological hypothesis was incorrect. As regards perceptual phonology, clearly a prompt is not applicable in this case.

Lee and Lyster (2016) note that many studies on phonological acquisition have not included corrective feedback in training because it was assumed that learner inferences would be based upon statistical learning, i.e. based upon exposure to input alone. This has been borne out in a study by Felker and colleagues (2021). In the study "listeners receiving contrastive corrective feedback began to accommodate their interlocutor’s accent over time" facilitating a speaker-specific vowel shift (Felker et al., 2021, p. 1047). These findings suggest that vowel perception has at least some malleability: learners can substitute phonemic categories or categorizations according to speakers is feedback is provided, although receiving feedback on how a categorization is incorrect appears to be necessary. Additionally, due to the feedback-induced malleability of the phonemes, it appears that although vowels may be perceived incorrectly initially, intended perceptions can be facilitated through feedback, which also appears to be supported by the work on high variability phonetic training (HVPT) such as that by Cebrian et al. (2019).

CALL Feedback

Regarding the timing of feedback, the question of what works needs to be accompanied by the the question of for whom? For example, Wang and Young (2012) found that adult learners have preferences for CALL feedback that does not interrupt learning activities (i.e. delayed feedback), but that younger learners have different preferences for feedback. On the other hand, while learner preferences should be taken into account, Wang and Young's study examines only

feedback preferences, not feedback efficacy; as such, investigation is needed to understand whether such learner preferences align with what is pedagogically effective.

Regarding aural language, Canals et al. (2020) found that a useful aspect for learners of delayed audio feedback on a speaking task was that they were able to review it several times. The advantages of review are that attention is given to the feedback, which allows learners to 'notice' (Schmidt, 1990) the difference between their incorrect answer and the intended correct answer. This may also be the case for other modalities, and also, potentially for listening instruction.

According to Cardenas Claros (2020), there is a predominance in CALL of feedback falling into a binary of correct and incorrect answers. Yet outside of language teaching and learning, Johnson and Maraffino (2021) state that feedback of correct/incorrect with presentation of the correct answer is also prevalent, and that "the most common forms include knowledge of results (KR) and knowledge of correct response (KCR)" (Ch. 34. p. 3/61). Obviously these modes of feedback being so prevalent is due to computational limitations. Healy and Warschauer (1998) advocate for applications that provide more detailed feedback that whether or not an answer is correct but for adaptive software that uses pattern recognition algorithms to respond "why it was right or wrong and offering suggestions for further study-going on to a more advanced level or doing some extra work at the current or a previous level" (p. 66). While the use of adaptive software is a conceptually interesting idea, its use in listening seems to be akin to the adage of using a sledgehammer to crack a nut.

Cognitive load (Sweller et al., 2011) dictates that only a limited amount of attention can be given to listening and/or phonological decoding. As learners engage mental facilities toward cognitive tasks, such as parsing phonetic information to phonemes, and phonemes to lexis, and lexis to semantic messages, less and less capacity is available for extraneous purposes. However if feedback given is unrelated to immediate learning objectives, it can potentially be counterproductive to the learning process because learners need to direct attention to the unnecessary information (Johnson & Maraffino, 2021). With these conditions in mind, the importance of not only salient input but also salient, relevant feedback is important in order to have learners process language and the feedback in the most effective way to facilitate acquisition.

Based on research carried out with L2 writing tasks, Rassaei (2019) recommends that "language teachers provide asynchronous audio-based and text-based CF on learners' out of class assignments in addition to in class face-to-face CF, given the flexible opportunities in terms of time and location that asynchronous text-based and audio-based CF can provide for L2 learning" (p. 108). Flexibility in time and location may not always be positive, because learners can choose to attempt to attend to feedback in suboptimal conditions, but equally classrooms can also be noisy, distracting places which may not always afford sufficient attention to either computer-facilitated feedback or face-to-face teacher feedback.

Fiorella, Vogel-Walcutt and Schatz (2012) suggest that feedback in a spoken modality for simulation-based training tasks results in better decision making than feedback provided by orthographic means. It must be noted that this is related to military training and not language learning, although the cognitive aspects of the feedback may be worth considering. Immediate teacher corrective feedback on pronunciation during L2 speaking tasks provided learners with the means to acquire more perceptual acuity of the /i-I/ contrast. (Zhang & Liao, 2023). Thus, the feedback on learners' own phonological hypotheses can provide input which leads to phonological acquisition. However, regarding the current study, when learning activities are provided through computers or mobile devices, it is not always possible to provide immediate face-to-face feedback.

RESEARCH QUESTIONS

The potential for vowel acquisition from feedback has not been fully explored. As such, this study explores the following research question:
Does immediate verbal feedback result in higher gains in perceptual vowel acquisition?

The preliminary hypothesis is that immediate verbal feedback results in higher gains in perceptual vowel acquisition. Additionally, the SLA studies generally point to immediate feedback providing greater efficiency to the acquisition process. However, it must be noted that none of the above studies have investigated vowel perception or identification, with the exception of, Zhang and Liao (2023).

METHODOLOGY

At a university English Medium Instruction programme in Japan, an intact class of 37 students at approximately CEFR B1-B2 level participated in a quasi-experiment consisting of a pretest, four lessons, and a post-test. Originally a delayed post-test was planned but this was abandoned due participant attrition across lessons and tests due to influenza and coronavirus infections. All tests and lessons were provided in a Moodle installation on a server run by the author, and were conducted in the classroom during a course subject to academic credit . If participants missed a lesson, they were encouraged to complete the lesson outside of class time.

Participants

As stated above, the participants come from an intact class of 37, from whom cooperation was elicited. All learners returned consent forms after the study was described in English and being given information sheets in English and Japanese. The activities were integrated into the classes and academic credit was awarded for participation.

Additionally, in using an intact class, ecological validity is maintained, therefore there is internal validity in using a sample drawn from the same community, and moreover, running the two groups at the same time with the same teacher-researcher, allows for factors such as time of day, instructor and classroom environment to be discounted from the data analysis.

From this class, 34 participants were kept. Three participants were removed due to failing to complete either the pretest or post-test. Participants were grouped based on pretest score by pairing similar score across groups, with 17 participants assigned to the group receiving immediate feedback (Immediate) and 17 participants assigned to the group receiving delayed feedback.

Tests

Tests consisted of 32 vowel identification items, eight for each vowel /æ/, /ʌ/, /ɜ:/, and /ɔ:/. Participants listened to each item and then chose the vowel identified. Vowels were not indicated by standard orthography or IPA symbols but by using photographs of iconic images: a cat for /æ/, the Sun for /ʌ/, a bird for /ɜ:/ and a door for /ɔ:/. Participants could listen to each item as many times as they wished, though they were encouraged to choose the answer they were most confident about if they were unsure.

Test materials can be accessed from https://zenodo.org/records/11025576

Lessons

The materials for each lesson were YouTube videos, listed in the Appendix section. These were chosen in relation to the themes of the English for Academic Purposes lessons taught by the author to the participants. Each lesson consisted of four rounds of vowel identification followed by an utterance transcription featuring the vowel in context. The utterance transcription served to benefit the participants by providing a listening task. After the four rounds of vowel identification and utterance transcription, participants summarized the edited video.

The immediate feedback group received recorded verbal feedback after completing each question in the lesson. The feedback provided was given in relation to pattern recognition algorithms that the Moodle platform allows using RegEx, a type of plaintext input frequently used in web programming and user interfaces for web software. The delayed feedback group were provided with verbal feedback on all correct answers after completing the four rounds of vowel identification and utterance transcription and video summary.

RESULTS

The raw results show that the gains overall for the Delayed group were higher than those of the Immediate group. Both groups made losses with the STRUT vowel, and the differences between groups are particularly wide for TRAP. However, the Immediate group made higher gains than the Delayed group with NURSE.

Table 1: Mean pretest to post-test gains of target vowels from immediate and delayed feedback groups

Group	Mean gain	Mean TRAP gain	Mean STRUT gain	Mean NURSE gain	Mean THOUGHT gain
Delayed	1.56	1.12	-0.188	0.125	0.5
Immediate	-0.412	-0.765	-0.647	0.353	-0.0588

Paired-sample t-tests were used to analyse the difference in feedback modalities for all of the vowels and also overall gains. The t-scores, generally have low values overall with correspondingly weak Bayes Factors, mainly at less than anecdotal strength probability. However, the t-score for TRAP gains suggests that delayed feedback is more useful and this also has a convincing Bayes Factor. As such, the t-test evidence shows that for most vowels, delayed feedback is likely to result in higher gains, particularly TRAP; however, the overall gains Bayes Factor is highly inconclusive, which suggests that the NURSE gains t-score skews the overall pattern. NURSE is the only vowel with a positive t-score, although the Bayes Factor is only at approximately anecdotal level.

Table 2: t-tests of pretest to post-test gains from immediate and delayed feedback groups

Description	t-score	Bayes Factor
Overall gains	-1.65	0.919
TRAP gains	-3.17	11.6
STRUT gains	-0.733	0.407
THOUGHT gains	-1.21	0.579
NURSE gains	0.435	0.357

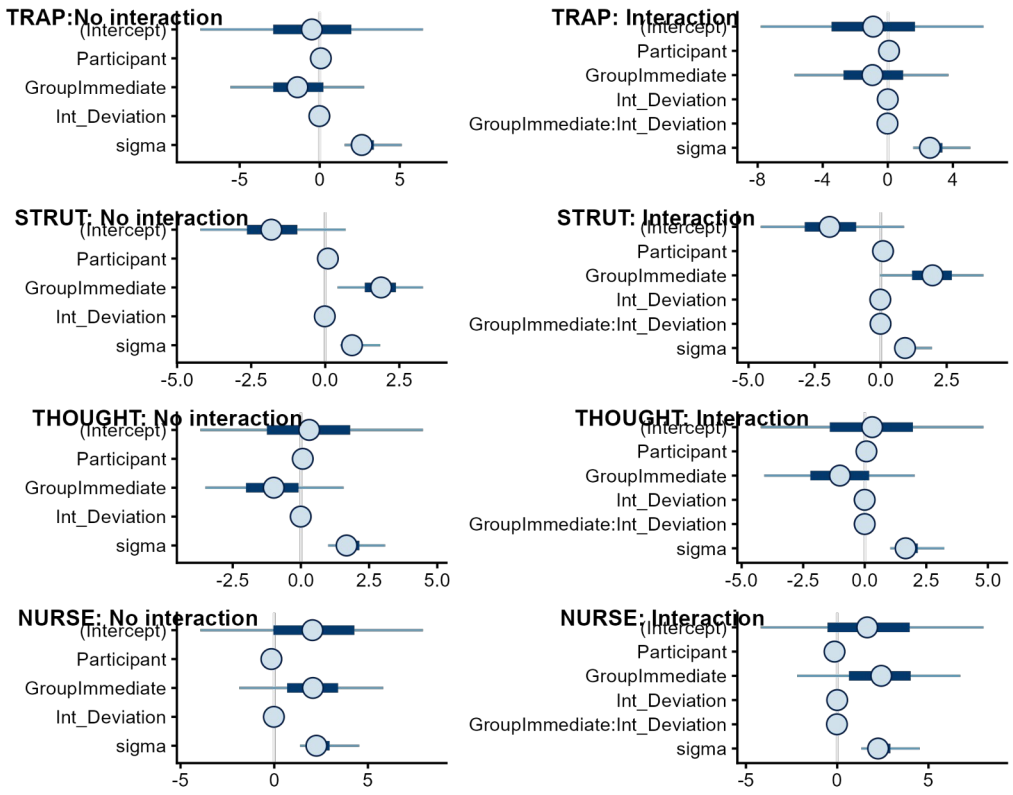
The use of a general linear mixed model (GLMM) allows for the detection of weights of factors contributing to the gains in acquisition. The variables tested for are participant, group, mean lesson score, standard deviation of lesson score, and standard deviation of interval between lessons and tests. All Bayes Factors are informative where figures are given. Lesson intervals cannot be said to have a direct influence upon gains, yet the standard deviation of intervals, this results in a Bayes Factor that suggests strong evidence for interaction (BF =2.634 x10^-4). However, a similarly strong Bayes Factor in the opposite direction (32.2 at 3 s.f.) was found for gains, condition, participant and individual vowel gains, suggesting no interaction between the factors. Additionally, when lesson score was entered into the model, only NA values were obtained. However, for individual vowels, condition, participant and interval standard deviation,

strongly informative Bayes factors were obtained (all < 0.2); however, again they suggest moderate evidence for no interaction between those factors.

Table 3: GLMMs and Bayes Factors for pretest-posttest gains

Descriptions	BF
Gains, condition, means and sds of lesson score: no interaction vs interaction	NA
Gains, condition, and sds of lesson score: no interaction vs interaction	0.0002634
Gains, condition, sd of lesson intervals: no interaction vs interaction	NA
Gains, condition, participant and individual vowel gains: no interaction vs interaction	32.1832319
TRAP gains, condition, participant and lesson mean and standard deviation: no interaction vs interaction	NA
STRUT gains, condition, participant and lesson mean and standard deviation: no interaction vs interaction	NA
THOUGHT gains, condition, participant and lesson mean and standard deviation: no interaction vs interaction	NA
NURSE gains, condition, participant and lesson mean and standard deviation: no interaction vs interaction	NA
TRAP gains, condition, participant and interval standard deviation: no interaction vs interaction	0.1751258
STRUT gains, condition, participant and interval standard deviation: no interaction vs interaction	0.0613003
THOUGHT gains, condition, participant and interval standard deviation: no interaction vs interaction	0.1410329
NURSE gains, condition, participant and interval standard deviation: no interaction vs interaction	0.1332181

Figure 1: Plots of GLMMS for vowel gains as a function of participant, group and interval standard deviation.



The standard deviation of interval between lessons effect on gains can be seen in Figure 1, where Int_Deviation is at around 0 on the GLMM plots. However, Group factors (seen in GroupImmediate) play a major part in the GLMM therefore there is an effect on the results, with GroupImmediate values between approximately -1 and 2.5 but notably not at 0.

CONCLUSION

Based on the results above, it can be concluded that delayed feedback is more likely than immediate feedback to result in perceptual acquisition of L2+ English vowels, as observed in the Group effects in the GLMMs. Furthermore, given the lack of interaction in the GLMMs between Group effects and the standard deviation of intervals between each participant's lesson completion, and the lack of effect in the t-tests, feedback timing is likely to have some effect, despite lower gains (or even losses for the immediate group) if learners cram, or if they are absent.

However, the standard deviation of interval between lessons effect on gains for individual vowels can be said to have a moderate evidence base against interaction, with Bayes Factors well below the ceiling threshold of 1/3 as recommended by Norouzian et al. (2019). Therefore, while the feedback timing of groups did not have a sufficiently large effect upon t-scores, the Group factor did affect the GLMMs, much more than the standard deviation of lesson intervals.

Regarding individual vowels, it may be seen that gains were readily observed in TRAP (/æ/) and THOUGHT (/ɔ:/) differences, whereas a very small between-groups difference was observed for NURSE (/ɜ:/) and losses for both groups with STRUT (/ʌ/). Due to the more peripheral positions in the vowel space of TRAP and THOUGHT, and the more central positions of STRUT and NURSE, it may be concluded that the Natural Referent Vowel framework (Polka & Bohn, 2011) is supported, albeit tentatively due to the relatively small group size and the attendance issues during the study period.

DISCUSSION

The results in this study are important for teachers and potentially also TELL materials developers. Feedback modality timing is important and can contribute to rate of acquisition, however it is important to acknowledge classroom realities: learner absences and failure to make up missed work, or not sufficiently spacing missed work, result in cramming. Such cramming of learning activities can turn planned, pedagogically-sound activities into unsound learning practices. The results and conclusion therefore must be interpreted on a case-by-case basis due to potential differences between teaching and learning contexts. However, ecologically valid results can contribute to the development of knowledge in the SLA and language teaching fields, less through an approach combining findings from several contexts, which may allow for consideration of how and when feedback timing has an effect, and with which types of populations.

In the advent of a global pandemics, absence from class and incomplete hyflex learning activities is likely to be an increasing experience for all educators including language teachers. Absence, is of course, a factor in language learning, because it results in a lack of exposure to the target language, whether in-person, remote synchronous or asynchronous as in distance and blended learning. Furthermore, even if learners make up incomplete tasks, there may be cramming effects,

resulting in low gains. In sum, disturbed and/or disrupted exposure to a language is detrimental to its acquisition. However, more research is necessary, and therefore encouraged, in order to investigate the size of any effects, in more usual class attendance situations.

Limitations

Participant attendance greatly affects just how generalizable the findings are across different populations. However, attendance in class cannot be guaranteed, especially when COVID isolation is considered. A further limitation of the current study is that a delayed post-test could not feasibly be carried out. Participant attrition can be a serious issue when working with samples from intact classes, and this is compounded by pandemic situations, due to difficulties in attendance due to infection with a highly contagious disease. The current study was conducted during the pandemic, at a time when in-class teaching with hybrid teaching for those unable to attend was encouraged. However, COVID-19 infection - even when not lethal - can be debilitating, and certainly in 'mild' cases leaves sufferers fatigued. As such, participants infected with COVID were unable, or understandably unmotivated to complete language learning exercises. Additionally, some language learners are more consistent with attendance than others for myriad reasons, which can, again be a factor in participant attrition with intact classes, particularly for completing testing sessions. Furthermore, due to participant attrition, a large sample size could not be maintained. While Bayesian methods are used to mitigate against small sample size (due to the central limit theorem not applying to Bayesian statistics), a larger sample may have resulted in more granularity in the data, thus being more informative. However, the ecological validity of the study being conducted in a classroom means that the transferability of findings is possible, and that the methodology can be repeated across contexts and extended further.

References

Archangeli, DB (1999) 'Introducing Optimality Theory', *Annual Review of Anthropology*, 28(1), pp. 531–552. Available at: https://doi.org/10.1146/annurev.anthro.28.1.531.

Baese-Berk, MM (2019) 'Interactions between speech perception and production during learning of novel phonemic categories', *Attention, Perception, & Psychophysics*, 81(4), pp. 981–1005. Available at: https://doi.org/10.3758/s13414-019-01725-4.

Baker, W *et al.* (2002) 'The Effect of Perceived Phonetic Similarity on Non-Native Sound Learning by Children and Adults', in *Proceedings of the 26th annual Boston University Conference on Language Development*. *the 26th annual Boston University Conference on Language Development*, Somerville, MA: Cascadilla Press, pp. 36–47. Available at: https://www.paveltrofimovich.ca/wp-content/uploads/2021/07/Baker_et_al_2002.pdf.

Best, CT (1995) 'A direct realist view of cross-language speech perception', in W. Strange (ed.) *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, pp. 171–206.

Best, CT and Tyler, MD (2007) 'Nonnative and second-language speech perception: Commonalities and complementarities', in O.-S. Bohn and M.J. Munro (eds) *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company (Language Learning & Language Teaching), pp. 13–34. Available at: https://doi.org/10.1075/lllt.17.07bes.

Bundgaard-Nielsen, RL *et al.* (2012) 'Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility', *Applied Psycholinguistics*, 33(3), pp. 643–664. Available at: https://doi.org/10.1017/S0142716411000518.

Bundgaard-Nielsen, RL, Best, CT and Tyler, MD (2011) 'Vocabulary Size is Associated with Second-Language Vowel Perception Performance in Adult Learners', *Studies in Second Language Acquisition*, 33(3), pp. 433–461. Available at: https://doi.org/10.1017/S0272263111000040.

Canals, L *et al.* (2020) 'Second Language Learners' and Teachers' Perceptions of Delayed Immediate Corrective Feedback in an Asynchronous Online Setting An Exploratory Study', *TESL Canada Journal*, 37(2), pp. 181–209. Available at: https://doi.org/10.18806/tesl.v37i2.1336.

Cárdenas-Claros, MS (2020) 'Conceptualizing feedback in computer-based L2 language listening', *Computer Assisted Language Learning*, 35(5–6), pp. 1168–1193. Available at: https://doi.org/10.1080/09588221.2020.1774615.

Crain, S, Koring, L and Thornton, R (2017) 'Language acquisition from a biolinguistic perspective', *Neuroscience & Biobehavioral Reviews*, 81, pp. 120–149. Available at: https://doi.org/10.1016/j.neubiorev.2016.09.004.

Darcy, I *et al.* (2012) 'Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English– L2 French acquisition', *Second Language Research*, 28(1), pp. 5–40. Available at: https://doi.org/10.1177/0267658311423455.

Escudero, P (2005) *Lingustic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. LOT.

Escudero, P and Williams, D (2014) 'Distributional learning has immediate and long-lasting effects', *Cognition*, 133(2), pp. 408–413. Available at: https://doi.org/10.1016/j.cognition.2014.07.002.

Faris, MM, Best, CT and Tyler, MD (2018) 'Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories', *Journal of Phonetics*, 70, pp. 1–19. Available at: https://doi.org/10.1016/j.wocn.2018.05.003.

Felker, E., Broersma, M and Ernestus, M (2021) 'The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue', *Applied Psycholinguistics*, 42(4), pp. 1029–1055. Available at: https://doi.org/10.1017/s0142716421000205.

Felker, E, Ernestus, M and Broersma, M (2019) 'Lexically Guided Perceptual Learning of a Vowel Shift in an Interactive L2 Listening Context', in *Interspeech 2019*. *Interspeech 2019*, ISCA, pp. 3123–3127. Available at: https://doi.org/10.21437/Interspeech.2019-1414.

Fiorella, L, Vogel-Walcutt, J and Schatz, S (2012) 'Applying the modality principle to real-time feedback and the acquisition of higher-order cognitive skills', *Educational Technology Research and Development*, 60, pp. 223–238. Available at: https://doi.org/10.1007/s11423-011-9218-1.

Flege, J. and Bohn, O-S (2021) 'The revised Speech Learning Model', in R. Wayland, (ed.) *Second language speech learning: Theoretical and empirical progress*. Cambridge University Press, pp. 3–83. Available at: https://doi.org/10.1017/9781108886901.002.

Flege, JE (1995) 'Second-language Speech Learning: Theory, Findings, and Problems.', in W. Strange (ed.) *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, pp. 229–273.

Flege, JE and MacKay, IRA (2004) 'Perceiving vowels in a second language', *Studies in Second Language Acquisition*, 26(01), pp. 1–34. Available at: https://doi.org/10.1017/S0272263104026117.

Georgiou, GP (2021) 'Toward a new model for speech perception: the Universal Perceptual Model (UPM) of second language', *Cognitive Processing* [Preprint]. Available at: https://doi.org/10.1007/s10339-021-01017-6.

Hardison, D (2017) 'Effects of contextual and visual cues on spoken language processing enhancing L2 perceptual salience through focused training', in S.M. Gass, P. Spinner, and J. Behney (eds) *Salience in Second Language Acquisition*. London: Routledge.

Ingram, JC and Park, S-G (1997) 'Cross-language vowel perception and production by Japanese and Korean learners of English', *Journal of Phonetics*, 25(3), pp. 343–370. Available at: https://doi.org/10.1006/jpho.1997.0048.

Johnson, CI and Maraffino, MD (2021) 'The feedback principle in multimedia learning', in Mayer, Richard E. and Fiorella, Logan (eds) *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge University Press, p. 34.1-34.64. Available at: https://doi.org/10.1017/9781108894333.

Lyster, R. and Ranta, L. (2013) 'Counterpoint piece: the case for variety in corrective feedback research', *Studies in Second Language Acquisition*, 35(1), pp. 167–184. Available at: https://doi.org/10.1017/S027226311200071X.

Nelson, C. (2022) 'Do a learner's background languages change with increasing exposure to L3? Comparing the multilingual phonological development of adolescents and adults', *Languages*, 7(78), p. 1-28. Available at: https://doi.org/10.3390/languages7020078.

Norouzian, R, de Miranda, M and Plonsky, L (2019) 'A Bayesian Approach to Measuring Evidence in L2 Research: An Empirical Investigation', *The Modern Language Journal*, 103(1), pp. 248–261. Available at: https://doi.org/10.1111/modl.12543.

Pierrehumbert, JB (2003) 'Phonetic diversity, statistical learning and acquistion of phonology', *Language & Speech*, 46(2/3), pp. 115–154. Available at: https://doi.org/10.1177/00238309030460020501.

Prince, A and Smolensky, P (2002) *Constraint Interaction in Generative Grammar*. Rutgers University, p. 262.

Rassaei, E (2019) 'Computer-mediated text-based and audio-based corrective feedback, perceptual style and L2 development', *System*, 82, pp. 97–110. Available at: https://doi.org/10.1016/j.system.2019.03.004.

Rebuck, M (2008) 'The effect of excessively difficult listening lessons on motivation and the influence of authentic listening as a "lesson-selling" tag', *JALT Journal*, 30(2), p. 197. Available at: https://doi.org/10.37546/JALTJJ30.2-3.

Rukthong, . and Brunfaut, T (2020) 'Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks', *Language Testing*, 37(1), pp. 31–53. Available at: https://doi.org/10.1177/0265532219871470.

Schmidt, R.W. (1990) 'The role of consciousness in second language learning', *Applied Linguistics*, 11(2), pp. 129–158. Available at: https://doi.org/10.1093/applin/11.2.129.

Sweller, J, Ayres, P and Kalyuga, S (2011) *Cognitive Load Theory*. New York, NY: Springer New York. Available at: https://doi.org/10.1007/978-1-4419-8126-4.

Tyler, MD (2019) 'PAM-L2 and phonological category acquisition in the foreign language classroom', in A.M. Nyvad et al. (eds) *A sound approach to language matters: in honor of Ocke-Schwen Bohn*. Aarhus University, pp. 607–630.

Vafaee, P. and Suzuki, Y. (2020) 'The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability', *Studies in Second Language Acquisition*, 42(2), pp. 383–410. Available at: https://doi.org/10.1017/S0272263119000676.

Van Leussen, J-W and Escudero, P (2015) 'Learning to perceive and recognize a second language: the L2LP model revised', *Frontiers in Psychology*, 6. Available at: https://doi.org/10.3389/fpsyg.2015.01000.

Warschauer, M and Healey, D (1998) 'Computers and language learning: an overview', *Language Teaching*, 31, pp. 57–71.

Winn, MB and Teece, KH (2021) 'Listening Effort Is Not the Same as Speech Intelligibility Score', *Trends in Hearing*, 25, p. 23312165211027688. Available at: https://doi.org/10.1177/23312165211027688.

Zhang, W and Liao, Y (2023) 'The role of auditory processing in L2 vowel learning: evidence from recasts', *Humanities and Social Sciences Communications*, 10(1), pp. 1–14. Available at: https://doi.org/10.1057/s41599-023-02001-5.

Appendix: YouTube video list

Lesson 1

IntuitInc. (2012, September 28). Lean Startup Lessons: Test Before you Build [Video]. Youtube. https://www.youtube.com/watch?v=KqtVWzwIfso

Lesson 2

Startup Amsterdam. (). # How to give the perfect pitch - with TedX speech coach David Beckett - Young Creators Summit 2016 [Video]. Youtube. https://www.youtube.com/watch?v=Njh3rKoGKBo

Lesson 3

Dartmouth. (2018, August 29). How Playing Games Can Change the World [Video]. Youtube. https://www.youtube.com/watch?v=h0u4accv15c

Lesson 4

Wired. (2014, December 16). Wired by design: a game designer explains the counterintuitive secret to fun [Video]. Youtube. https://www.youtube.com/watch?v=78rPt0RsosQ

Appendix 6: Data and scripts relating to Jones and Blume (2022) [Appendix 1]

Appendix 6.1: Lesson data

Data x2a.csv

Participant	Condition	Pre score	L1	L2	L3	L4	L5	Post date	Post score
1	N	28	04/12/2021 11:00	06/12/2021 14:24	27/12/2021 11:18	27/12/2021 12:28	28/12/2021 16:01	22/01/2022 13:27	21
2	G	28	27/01/2022 23:44	27/01/2022 23:55	28/01/2022 00:05	28/01/2022 00:15	28/01/2022 00:25	28/01/2022 00:33	23
3	G	26	27/01/2022 23:44	27/01/2022 23:55	28/01/2022 00:05	28/01/2022 00:13	28/01/2022 00:24	28/01/2022 00:33	25
4	G	27	24/12/2021 22:39	07/01/2022 00:00	15/01/2022 20:04	18/01/2022 01:26	25/01/2022 21:56	25/01/2022 22:17	22
5	N	30	27/01/2022 17:36	27/01/2022 18:05	27/01/2022 18:26	27/01/2022 19:14	27/01/2022 19:25	27/01/2022 19:32	14
6	N	29	14/12/2021 11:48	14/12/2021 12:12	14/12/2021 20:56	15/12/2021 10:49	24/12/2021 19:22	24/12/2021 19:36	29
7	N	28	21/01/2022 02:11	28/01/2022 20:37	28/01/2022 21:15	28/01/2022 21:51	28/01/2022 22:27	28/01/2022 22:45	27
8	G	30	06/12/2021 09:38	09/12/2021 11:56	09/12/2021 15:34	19/12/2021 20:54	25/12/2021 11:23	20/01/2022 09:58	29
9	N	26	02/12/2021 00:06	05/12/2021 14:42	11/12/2021 00:12	24/12/2021 16:53	24/12/2021 17:32	20/01/2022 11:15	24
10	G	28	15/12/2021 10:23	15/12/2021 10:46	15/12/2021 15:19	17/12/2021 21:31	13/01/2022 07:08	13/01/2022 11:05	24
11	N	21	18/01/2022 00:00	18/01/2022 00:23	18/01/2022 00:43	18/01/2022 01:42	18/01/2022 05:48	18/01/2022 05:58	22
12	G	27	01/12/2021 23:40	26/01/2022 23:12	26/01/2022 23:38	27/01/2022 00:36	27/01/2022 00:56	27/01/2022 09:52	26
13	N	29	26/11/2021 21:41	05/12/2021 14:42	30/12/2021 21:46	31/12/2021 16:13	22/01/2022 21:46	24/01/2022 15:12	29
14	G	29	01/12/2021 22:41	05/12/2021 18:04	12/12/2021 11:19	19/12/2021 11:09	03/01/2022 11:14	15/01/2022 14:00	27
15	N	29	02/12/2021 20:53	15/12/2021 16:51	15/12/2021 17:07	15/12/2021 23:33	22/12/2021 23:10	12/01/2022 22:43	29
16	G	17	29/11/2021 00:01	23/01/2022 00:50	25/01/2022 14:17	25/01/2022 15:24	25/01/2022 16:08	25/01/2022 16:39	20

Appendix 6.2: Pretest data

X2apretest

Participan	Time taken	Grade/32.	BT10	BT11	BT12	BT13	BT14	BT15	BT16	BT17	BT18	BT19	BT1	BT2	BT3	BT4
1	1 day 4 hours	28	0	1	1	1	1	1	1	1	1	1	0	1	1	1
2	1 hour 31 mins	28	0	1	1	0	1	1	1	1	1	1	0	1	1	1
3	1 hour 23 mins	26	0	1	1	0	0	1	0	1	1	1	1	0	1	1
4	1 hour 13 mins	27	1	1	1	1	1	1	0	1	0	1	0	1	1	1
5	50 mins 21 secs	30	1	1	1	1	1	1	1	1	1	1	0	1	1	1
6	52 mins 55 secs	29	1	1	1	1	0	1	1	1	1	1	0	1	1	1
7	1 hour	27	1	1	1	0	1	1	1	1	1	0	0	1	0	1
8	51 mins 16 secs	30	1	1	1	1	1	1	1	1	0	1	1	1	1	1
9	53 mins 52 secs	26	1	1	0	1	1	1	1	1	1	0	0	1	1	1
10	1 hour 10 mins	28	0	1	1	1	1	1	0	1	1	1	0	1	1	1
11	8 hours 17 mins	21	0	1	1	0	0	1	1	1	1	0	0	1	0	1
12	5 hours 48 mins	27	1	1	1	0	1	1	1	1	1	1	0	1	1	1
13	40 mins 54 secs	29	1	0	1	1	1	1	1	1	1	1	0	1	1	1
14	56 mins 6 secs	29	1	0	1	1	1	1	1	1	1	1	0	1	1	1
15	1 hour 24 mins	29	0	1	1	1	1	1	1	1	1	1	0	1	1	1
16	1 day 10 hours	17	0	1	1	0	0	0	0	1	0	0	1	1	0	1

Participan	BT5	BT6	BT7	BT8	BT9	BT20	BT21	BT22	BT23	BT24	BT25	BT26	BT27	BT28	BT29	BT30	BT31	BT32
1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1
2	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
6	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
9	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1
10	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	1
12	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	1
13	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	0	1

Appendix 6.3: Post test data

X2A Post test

Participant	Time taken	Grade/32.	BT10	BT11	BT12	BT13	BT14	BT15	BT16	BT17	BT18	BT19	BT1	BT2	BT3	BT4
1	17 mins 40 secs	21	0	1	1	0	1	1	0	1	1	1	0	1	0	0
2	7 mins 35 secs	23	1	1	1	1	1	1	0	1	1	0	1	1	0	1
3	8 mins 40 secs	25	0	1	1	0	1	1	1	1	1	1	0	1	0	1
4	20 mins 43 secs	22	1	1	1	0	1	0	0	1	1	1	0	0	1	1
5	7 mins 10 secs	14	0	0	1	1	0	0	0	0	1	1	0	0	0	0
6	13 mins 22 secs	29	0	1	1	1	1	1	1	1	1	0	0	1	1	1
7	17 mins 14 secs	28	0	1	1	0	1	1	1	1	1	1	1	1	0	1
8	8 mins 16 secs	29	1	1	1	1	1	1	1	1	1	0	1	1	1	1
9	16 mins 37 secs	24	1	1	1	0	0	1	1	1	0	1	0	1	0	1
10	7 mins 54 secs	24	1	0	1	1	0	1	1	1	1	1	0	1	0	0
11	10 mins 13 secs	22	1	1	1	1	1	1	0	0	1	1	0	1	0	0
12	16 mins 18 secs	26	1	1	1	0	1	1	0	1	0	1	0	1	1	1
13	8 mins 43 secs	29	1	1	1	1	1	0	1	1	1	1	0	1	1	1
14	10 mins 22 secs	27	0	1	1	0	0	1	0	1	1	1	1	1	1	1
15	8 mins 19 secs	29	1	1	1	1	1	1	1	1	1	0	1	1	1	1
16	30 mins 6 secs	20	1	1	1	0	0	1	0	1	1	0	0	1	0	1

Participan	BT5	BT6	BT7	BT8	BT9	BT20	BT21	BT22	BT23	BT24	BT25	BT26	BT27	BT28	BT29
1	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1
2	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0
3	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
4	1	0	1	1	0	1	1	1	0	1	1	0	1	0	0
5	0	1	1	0	0	0	1	1	1	0	0	1	0	1	0
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
7	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
8	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0
9	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
10	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0
11	1	0	1	1	1	1	0	1	1	1	0	1	0	0	0
12	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	0	1	1	1	1	0	1	1	1	0	1	0	0

Appendix 6.4: R script to analyse data in appendices 6.1-6.3, pertaining to appendix 1

x2ascript_3.r

```
#import data

x2a_data <- read.csv('x2a.csv')

x2apretest <- read.csv('x2apretest.csv')

x2aposttest <- read.csv('x2aposttest.csv')

# enables use of the same R packages I used when coding this data

library(groundhog)

# tabulate

groundhog.library("tidyverse", "2022-04-28", tolerate.R.version='4.0.3')

groundhog.library("lubridate", "2022-04-28", tolerate.R.version='4.0.3')

x2a_table <- tibble(x2a_data)

x2a_table <- mutate(x2a_table, L1 = parse_date_time(L1, orders = "%d/%m/%y %R"),

                    L2 = parse_date_time(L2, orders = "%d/%m/%y %R"), L3 =
parse_date_time(L3, orders = "%d/%m/%y %R"),

                    L4 = parse_date_time(L4, orders = "%d/%m/%y %R"), L5 =
parse_date_time(L5, orders = "%d/%m/%y %R"),

                    Post.date = parse_date_time(Post.date, orders = "%d/%m/%y %R"))

x2apre <- tibble(x2apretest)

x2apost <- tibble(x2aposttest)

#calculate scores for each vowel in both tests

x2apre <- mutate(x2apre, preTrap = (BT5 + BT6 + BT8 + BT10 + BT18 + BT21 + BT25 +
BT32))

x2apre <- mutate(x2apre, preStrut = (BT1+ BT4+ BT12+ BT16+ BT17+ BT26+ BT27+
BT28))
```

```
x2apre <- mutate(x2apre, preThought = (BT2 +  BT9 +  BT11 + BT15 +  BT20 +  BT23 +
BT24 +  BT29))

x2apre <- mutate(x2apre, preNurse = (BT3 +  BT7 +  BT13 +  BT14 +  BT19 +  BT22 +
BT30 +  BT31))

x2apost <- mutate(x2apost, postTrap = (BT5 + BT6 + BT8 + BT10 + BT18 + BT21 +
BT25 + BT32))

x2apost <- mutate(x2apost, postStrut = (BT1+ BT4+ BT12+ BT16+ BT17+ BT26+ BT27+
BT28))

x2apost <- mutate(x2apost, postThought = (BT2 +  BT9 +  BT11 + BT15 +  BT20 +  BT23
+  BT24 +  BT29))

x2apost <- mutate(x2apost, postNurse = (BT3 +  BT7 +  BT13 +  BT14 +  BT19 +  BT22 +
BT30 +  BT31))

#calculate times intervals between lessons/ post test in days

x2a_table <- mutate (x2a_table, L1_2int = as.numeric(difftime(L2, L1, units ="days")),

                    L2_3int = as.numeric(difftime(L3, L2, units ="days")),

                    L3_4int = as.numeric(difftime(L4, L3, units ="days")),

                    L4_5int = as.numeric(difftime(L5, L4, units ="days")),

                    L5_Pint = as.numeric(difftime(Post.date, L5, units ="days")))

#calculate standard deviation for intervals between lessons and posttest for each participant:
Ok 20220426

x2a_table2b <- select(x2a_table, L1_2int:L5_Pint)

x2a_table2b <- mutate(x2a_table2b, IntMean = summarise(rowwise(x2a_table2b),
mean(c_across(everything()))))

x2a_table2b <- mutate(x2a_table2b, Standard_Deviation = as.numeric(unlist

                                        (((L1_2int - IntMean)^2) + ((L2_3int -
IntMean)^2)

                                        + ((L3_4int - IntMean)^2) + ((L4_5int -
IntMean)^2)

                                        + ((L5_Pint - IntMean)^2) / 5)))

x2a_table <- mutate(x2a_table, Standard_Deviation = (x2a_table2b$Standard_Deviation))
```

```
# calculate gains

x2a_table <- mutate(x2a_table, gains = (Post.score - Pre.score))

x2a_table <- mutate(x2a_table, gainTrap = (x2apost$postTrap - x2apre$preTrap))

x2a_table <- mutate(x2a_table, gainStrut = (x2apost$postStrut - x2apre$preStrut))

x2a_table <- mutate(x2a_table, gainThought = (x2apost$postThought - x2apre$preThought))

x2a_table <- mutate(x2a_table, gainNurse = (x2apost$postNurse - x2apre$preNurse))

#Group tibble for summarising

x2a_table3 <- group_by(x2a_table, Condition)

x2atable3a <- summarize(x2a_table3, MeanPre = mean(Pre.score), MeanPost = mean(Post.score),

                MeanL1_2 = mean(L1_2int), MeanL2_3 = mean(L2_3int), MeanL3_4 = mean(L3_4int),

                MeanL4_5 = mean(L4_5int), MeanL5_Post = mean(L5_Pint), MeanIntSD = mean(Standard_Deviation))

# make linear model(s): OK 20220427

groundhog.library("rstan", "2022-04-28", tolerate.R.version='4.0.3')

groundhog.library("rstanarm", "2022-04-28", tolerate.R.version='4.0.3')

groundhog.library("bridgesampling", "2022-04-28", tolerate.R.version='4.0.3')

#Gains, condition and intervals no interaction vs interaction

model1 <- stan_glm(gains ~ Participant + Condition + L1_2int + L2_3int + L3_4int + L4_5int + L5_Pint,

                data = x2a_table, family = gaussian,

                diagnostic_file = file.path(tempdir(), "df.csv"))

model2 <- stan_glm(gains ~ Participant + Condition * (L1_2int + L2_3int + L3_4int + L4_5int + L5_Pint),

                data = x2a_table, family = gaussian,

                diagnostic_file = file.path(tempdir(), "df.csv"))
```

```
#Gains, condition, participant and interval standard deviation: no interaction vs interaction
model3 <- stan_glm(gains ~ Participant + Condition + Standard_Deviation,
                   data = x2a_table, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model4 <- stan_glm(gains ~ Participant + Condition * Standard_Deviation,
                   data = x2a_table, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
#Gains, participant and interval standard deviation: no interaction vs interaction
model5 <- stan_glm(gains ~ Participant + Standard_Deviation,
                   data = x2a_table, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model6 <- stan_glm(gains ~ Participant * Standard_Deviation,
                   data = x2a_table, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
#Gains, condition, participant and individual vowel gains: no interaction vs interaction
model7 <- stan_glm(gains ~ Participant + Condition + gainTrap + gainStrut + gainThought
+ gainNurse,
                   data = x2a_table, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model8 <- stan_glm(gains ~ Participant + Condition * (gainTrap + gainStrut + gainThought
+ gainNurse),
                   data = x2a_table, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
#Specific vowel gains, participant and interval standard deviation: no interaction vs
interaction
model9 <- stan_glm(gainTrap ~ Participant + Condition + Standard_Deviation,
                   data = x2a_table, family = gaussian,
```

```
                diagnostic_file = file.path(tempdir(), "df.csv"))
model10 <- stan_glm(gainTrap ~ Participant + Condition * Standard_Deviation,
                data = x2a_table, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model11 <- stan_glm(gainStrut ~ Participant + Condition + Standard_Deviation,
                data = x2a_table, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model12 <- stan_glm(gainStrut ~ Participant + Condition * Standard_Deviation,
                 data = x2a_table, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
model13 <- stan_glm(gainThought ~ Participant + Condition + Standard_Deviation,
                data = x2a_table, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model14 <- stan_glm(gainThought ~ Participant + Condition * Standard_Deviation,
                 data = x2a_table, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
model15 <- stan_glm(gainNurse ~ Participant + Condition + Standard_Deviation,
                data = x2a_table, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model16 <- stan_glm(gainNurse ~ Participant + Condition * Standard_Deviation,
                 data = x2a_table, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
#compare models to get a BF
mod1v2 <- bayes_factor(bridge_sampler(model1), bridge_sampler(model2))
mod3v4 <- bayes_factor(bridge_sampler(model3), bridge_sampler(model4))
```

```
mod5v6 <- bayes_factor(bridge_sampler(model5), bridge_sampler(model6))

mod7v8 <- bayes_factor(bridge_sampler(model7), bridge_sampler(model8))

mod9v10 <- bayes_factor(bridge_sampler(model9), bridge_sampler(model10))

mod11v12 <- bayes_factor(bridge_sampler(model11), bridge_sampler(model12))

mod13v14 <- bayes_factor(bridge_sampler(model13), bridge_sampler(model14))

mod15v16 <- bayes_factor(bridge_sampler(model15), bridge_sampler(model16))

allmods <- c(mod1v2$bf, mod3v4$bf, mod5v6$bf, mod7v8$bf, mod9v10$bf,
mod11v12$bf, mod13v14$bf, mod15v16$bf)

#Add to table

descs <- c("Gains, condition and intervals no interaction vs interaction",

        "Gains, condition, participant and interval standard deviation: no interaction vs
interaction",

        "Gains, participant and interval standard deviation: no interaction vs interaction",

        "Gains, condition, participant and individual vowel gains: no interaction vs
interaction",

        "TRAP gains, condition, participant and interval standard deviation: no interaction
vs interaction",

        "STRUT gains, condition, participant and interval standard deviation: no interaction
vs interaction",

        "THOUGHT gains, condition, participant and interval standard deviation: no
interaction vs interaction",

        "NURSE gains, condition, participant and interval standard deviation: no interaction
vs interaction")

modelTable <- data.frame(descs, allmods)

modelTibble <- modelTable %>% tibble()

groundhog.library("dplyr", "2022-04-28", tolerate.R.version='4.0.3')

modelTibble <- rename(modelTibble, Description = descs, "Bayes Factor" = allmods)

modelTibble
```

```
groundhog.library("BEST", "2022-04-28", tolerate.R.version='4.0.3')

#bayes-t-tests

#overall gains

groupG <- filter(x2a_table, Condition == "G")

groupP <- filter(x2a_table, Condition == "N")

tchainsAll <- BESTmcmc(groupG$gains, groupP$gains, priors = NULL, parallel =
FALSE)

plot(tchainsAll)

t_gainsAll <- t.test(groupG$gains, groupP$gains)

bayesfactorvalueAll <- bayes_factor(mean(groupP$gains), mean(groupG$gains))

meanGainDiffs <- mean(groupG$gains) - mean(groupP$gains)

#TRAP gains

tchainsTrap <- BESTmcmc(groupG$gainTrap, groupP$gainTrap, priors = NULL, parallel
= FALSE)

plot(tchainsTrap)

t_gainsTrap <- t.test(groupG$gainTrap, groupP$gainTrap)

bayesfactorvalueTrap <- bayes_factor(mean(groupP$gainTrap), mean(groupG$gainTrap))

meanTrapGainDiffs <- mean(groupG$gainTrap) - mean(groupP$gainTrap)

#STRUT gains

tchainsStrut <- BESTmcmc(groupG$gainStrut, groupP$gainStrut, priors = NULL, parallel
= FALSE)

plot(tchainsStrut)

t_gainsStrut <- t.test(groupG$gainStrut, groupP$gainStrut)

bayesfactorvalueStrut <- bayes_factor(mean(groupP$gainStrut), mean(groupG$gainStrut))

meanStrutGainDiffs <- mean(groupG$gainStrut) - mean(groupP$gainStrut)

#THOUGHT gains
```

```
tchainsThought <- BESTmcmc(groupG$gainThought, groupP$gainThought, priors =
NULL, parallel = FALSE)

plot(tchainsThought)

t_gainsThought <- t.test(groupG$gainThought, groupP$gainThought)

bayesfactorvalueThought <- bayes_factor(mean(groupP$gainThought),
mean(groupG$gainThought))

meanThoughtGainDiffs <- mean(groupG$gainThought) - mean(groupP$gainThought)

#NURSE gains

tchainsNurse <- BESTmcmc(groupG$gainNurse, groupP$gainNurse, priors = NULL,
parallel = FALSE)

plot(tchainsNurse)

t_gainsNurse <- t.test(groupG$gainNurse, groupP$gainNurse)

bayesfactorvalueNurse <- bayes_factor(mean(groupP$gainNurse),
mean(groupG$gainNurse))

meanNurseGainDiffs <- mean(groupG$gainNurse) - mean(groupP$gainNurse)

#Tabulate BFs for t-tests

GainDescs <- c("Overall gains", "TRAP gains", "STRUT gains", "THOUGHT gains",
"NURSE gains")

allts <- c(t_gainsAll$statistic, t_gainsTrap$statistic, t_gainsStrut$statistic,
t_gainsThought$statistic, t_gainsNurse$statistic)

allbfs <- c(bayesfactorvalueAll$bf, bayesfactorvalueTrap$bf, bayesfactorvalueStrut$bf,
bayesfactorvalueThought$bf, bayesfactorvalueNurse$bf)

gainsTable <- data.frame(GainDescs, allts, allbfs)

gainsTibble <- gainsTable %>% tibble()

gainsTibble <- rename(gainsTibble, Description = GainDescs)

gainsTibble <- rename(gainsTibble, 't Score' = allts)

gainsTibble <- rename(gainsTibble, 'BayesFactor' = allbfs)

#set up plots
```

```
gainsPlot <-  ggplot(x2a_table, aes(x = Condition, y = gains, fill = Condition)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  scale_fill_discrete(name = "Condition", labels = c("G", "N")) +
  theme(axis.ticks.x = element_blank()) +
  ggtitle("Overall gains")
trapPlot <-  ggplot(x2a_table, aes(x = Condition, y = gainTrap, fill = Condition)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("TRAP gains")
strutPlot <-  ggplot(x2a_table, aes(x = Condition, y = gainStrut, fill = Condition)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("STRUT gains")
thoughtPlot <- ggplot(x2a_table, aes(x = Condition, y = gainThought, fill = Condition)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("THOUGHT gains")
nursePlot <-  ggplot(x2a_table, aes(x = Condition, y = gainNurse, fill = Condition)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
```

```
  ggtitle("NURSE gains")

# Saves plots all together in a PNG file

groundhog.library("cowplot", "2022-04-28",tolerate.R.version='4.0.3')

plot_grid(gainsPlot, plot_grid(trapPlot, strutPlot), plot_grid(thoughtPlot, nursePlot), ncol =
1)

ggsave(filename="gains.png")
```

Appendix 7: Data and scripts pertaining to Jones (2024c) [Appendix 3]

Appendix 7.1: Lesson data

X1c_lessons_cleaned.csv

	Participan	Group	Pretest.to	Lesson.1.	Lesson.1.	Lesson.2.	Lesson.2.	Lesson.3.	Lesson.3.	Lesson.4.	Lesson.4.	Lesson.5.	Lesson.5.	Posttest.t	Delay.Tes	Lesson.M	Lesson.SI
1	3	AV	25	NA	55.56	NA	16.67	NA	55.56	NA	44.44	NA	55.56	25	NA	45.558	16.85146
2	19	A-Only	21	55.56	NA	55.56	NA	77.78	NA	77.78	NA	55.56	NA	28	29	64.448	12.1704
3	5	AV	20	NA	22.22	NA	44.44	NA	55.56	NA	11.11	NA	33.33	16	20	33.332	17.56961
4	21	A-Only	28	44.44	NA	55.56	NA	55.56	NA	33.33	NA	44.44	NA	29	27	46.666	9.300074
5	34	A-Only	NA	44.44	NA	22.22	NA	66.67	NA	NA	NA	55.56	NA	21	NA	47.2225	18.97882
6	31	A-Only	25	22.22	NA	33.33	NA	66.67	NA	44.44	NA	44.44	NA	22	NA	42.22	16.4825
7	8	A-Only	24	55.56	NA	66.67	NA	44.44	NA	55.56	NA	44.44	NA	29	NA	53.334	9.300074
8	26	AV	15	NA	NA	NA	11.11	NA	77.78	NA	NA	NA	NA	9	NA	44.445	47.14281
9	12	A-Only	22	44.44	NA	55.56	NA	66.67	NA	66.67	NA	33.33	NA	23	NA	53.334	14.49065
10	29	A-Only	15	22.22	NA	33.33	NA	55.56	NA	33.33	NA	NA	NA	15	NA	36.11	13.98441
11	35	AV	NA	NA	44.44	NA	33.33	NA	44.44	NA	55.56	NA	33.33	21	NA	42.22	9.298879
12	27	AV	17	NA	33.33	NA	44.44	NA	66.67	NA	33.33	NA	44.44	21	24	44.442	13.611
13	17	AV	30	NA	44.44	NA	33.33	NA	44.44	NA	44.44	NA	55.56	30	NA	44.442	7.859492
14	18	AV	22	NA	66.67	NA	33.33	NA	33.33	NA	55.56	NA	44.44	25	NA	46.666	14.49065
15	30	AV	19	NA	33.33	NA	44.44	NA	55.56	NA	22.22	NA	55.56	21	NA	42.222	14.49027
16	9	A-Only	17	44.44	NA	77.78	NA	44.44	NA	55.56	NA	55.56	NA	20	NA	55.556	13.611
17	28	AV	16	NA	33.33	NA	44.44	NA	44.44	NA	55.56	NA	22.22	17	18	39.998	12.67042
18	1	A-Only	20	44.44	NA	33.33	NA	NA	NA	44.44	NA	33.33	NA	20	NA	38.885	6.414361
19	33	AV	17	NA	11.11	NA	33.33	NA	55.56	NA	22.22	NA	55.56	24	NA	35.556	19.8792
20	32	AV	22	NA	22.22	NA	55.56	NA	55.56	NA	22.22	NA	33.33	22	NA	37.778	16.85442
21	14	AV	15	NA	33.33	NA	44.44	NA	44.44	NA	55.56	NA	33.33	14	NA	42.22	9.298879
22	4	A-Only	15	44.44	NA	66.67	NA	55.56	NA	44.44	NA	33.33	NA	22	18	48.888	12.67217
23	23	AV	24	NA	22.22	NA	66.67	NA	55.56	NA	33.33	NA	44.44	23	NA	44.444	17.5712
24	25	A-Only	24	NA	NA	NA	NA	NA	NA	NA	NA	NA	66.67	23	NA	66.67	NA
25	15	AV	23	NA	55.56	NA	33.33	NA	55.56	NA	55.56	NA	33.33	23	NA	46.668	12.17587
26	16	A-Only	15	33.33	NA	44.44	NA	NA	NA	33.33	NA	33.33	NA	14	16	36.1075	5.555
27	2	A-Only	25	66.67	NA	66.67	NA	88.89	NA	66.67	NA	55.56	NA	27	28	68.892	12.1704
28	7	A-Only	18	0	NA	44.44	NA	44.44	NA	33.33	NA	55.56	NA	19	NA	35.554	21.37285
29	11	AV	25	NA	33.33	NA	55.56	NA	22.22	NA	22.22	NA	44.44	23	22	35.554	14.48912
30	6	A-Only	25	33.33	NA	33.33	NA	55.56	NA	55.56	NA	44.44	NA	24	28	44.444	11.115
31	10	A-Only	23	55.56	NA	55.56	NA	66.67	NA	55.56	NA	44.44	NA	19	NA	55.558	7.859492
32	20	A-Only	16	22.22	NA	0	NA	66.67	NA	33.33	NA	11.11	NA	23	NA	26.666	25.58105
33	24	A-Only	24	33.33	NA	44.44	NA	55.56	NA	33.33	NA	44.44	NA	24	NA	42.22	9.298879
34	22	AV	21	NA	44.44	NA	44.44	NA	55.56	NA	44.44	NA	33.33	21	NA	44.442	7.859492
35	13	A-Only	20	33.33	NA	NA	NA	55.56	NA	44.44	NA	33.33	NA	19	NA	41.665	10.64137

Appendix 7.2: Pretest data

X1c_pretest_cleaned.csv

	Participan	Group	Started.on	Complete	Time.take	Score	Q..1.1.00	Q..2.1.00	Q..3.1.00	Q..4..1.00	Q..5..1.00	Q..6..1.00	Q..7..1.00	Q..8..1.00	Q..9..1.00	Q..10..1.0(	Q..11..1.0(	Q..12..1.0(	Q..13..1.0(	Q..14..1.0(	Q..15..1.0(	Q..16..1.0(
1	1	A-Only	#######	########	6 mins 39	20	1	1	1	0	1	1	1	1	0	0	0	1	0	1	1	1
2	2	A-Only	#######	########	12 mins 1	25	0	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1
3	3	AV	#######	########	10 mins 5	25	1	1	1	1	1	1	0	0	1	1	1	0	1	0	1	0
4	4	A-Only	#######	########	6 mins 17	15	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0
5	5	AV	#######	########	12 mins 5	20	0	1	1	0	0	1	1	1	1	0	0	1	0	0	1	1
6	6	A-Only	#######	########	11 mins 2	25	0	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1
7	7	A-Only	#######	########	13 mins 1(	18	0	1	1	1	1	0	1	1	0	0	1	1	0	0	1	0
8	8	A-Only	#######	########	9 mins 41	24	0	1	1	0	0	1	1	1	0	1	1	0	1	1	1	1
9	9	A-Only	#######	########	12 mins 5	17	0	1	1	0	1	1	0	0	0	0	0	1	1	0	0	1
10	10	A-Only	#######	########	22 mins 9	23	0	1	1	0	1	1	0	1	0	0	0	1	1	1	1	1
11	11	AV	#######	########	6 mins 38	25	0	1	1	1	0	0	0	1	1	1	1	1	0	1	0	1
12	12	A-Only	#######	########	20 mins 5	22	0	1	1	1	1	1	0	0	1	1	1	1	0	0	0	1
13	13	A-Only	#######	########	10 mins 5	20	0	1	0	1	1	1	0	0	1	1	0	1	0	0	1	1
14	14	AV	#######	########	19 mins 2	15	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1
15	15	AV	#######	########	17 mins 5	23	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1	1
16	16	A-Only	#######	########	10 mins 3	15	0	1	0	0	1	1	1	0	1	0	0	0	0	0	0	1
17	17	AV	#######	########	14 mins 3	30	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
18	18	AV	#######	########	18 mins 1	22	1	1	1	0	1	1	0	0	1	1	0	0	0	0	1	1
19	19	A-Only	#######	########	25 mins 3	21	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1
20	20	A-Only	#######	########	17 mins 2	16	0	1	1	0	0	0	0	0	1	0	0	1	1	0	1	1
21	21	A-Only	#######	########	17 mins 8	28	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0
22	22	AV	#######	########	12 mins 5	21	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1
23	23	AV	#######	########	15 mins 1	24	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	1
24	24	A-Only	#######	########	19 mins 3	24	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1
25	25	A-Only	#######	########	20 mins 5	24	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1	1
26	26	AV	#######	########	25 mins 5	15	1	1	1	1	1	1	0	0	1	0	0	0	0	0	1	1
27	27	AV	#######	########	22 mins 4	17	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1
28	28	AV	#######	########	19 mins 4	16	1	1	1	0	1	0	0	0	1	0	0	1	0	0	1	1
29	29	A-Only	#######	########	17 mins 9	15	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1
30	30	AV	#######	########	16 mins 4	19	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	1
31	31	A-Only	#######	########	13 mins 5	25	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1
32	32	AV	#######	########	12 mins 1	22	1	1	1	0	0	1	1	0	1	0	1	1	0	1	1	1
33	33	AV	#######	########	13 mins 2	17	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1	1

	Participan	Q..18..1.00	Q..19..1.00	Q..20..1.00	Q..21..1.00	Q..22..1.00	Q..23..1.00	Q..24..1.00	Q..25..1.00	Q..26..1.00	Q..27..1.00	Q..28..1.00	Q..29..1.00	Q..30..1.00	Q..31..1.00	Q..32..1.00
1	1	1	1	1	0	0	1	1	1	0	1	0	1	0	0	1
2	2	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1
3	3	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1
4	4	1	1	1	0	1	1	1	1	1	1	0	1	0	1	0
5	5	1	1	1	1	0	1	1	1	0	1	0	1	1	0	1
6	6	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1
7	7	0	1	1	1	1	1	1	0	0	1	0	0	0	1	1
8	8	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1
9	9	0	1	1	0	1	1	1	1	1	0	0	1	0	1	1
10	10	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1
11	11	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
12	12	0	1	1	0	1	1	1	0	1	0	1	1	1	1	1
13	13	0	1	1	1	0	1	1	1	0	1	0	1	0	1	1
14	14	0	1	1	1	0	1	1	1	1	0	0	1	0	0	1
15	15	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1
16	16	0	1	0	0	1	0	1	1	0	1	0	1	0	1	1
17	17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	18	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1
19	19	0	0	0	1	1	0	1	1	1	1	0	1	0	1	1
20	20	0	1	1	0	1	1	1	0	0	0	1	1	0	1	0
21	21	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
22	22	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1
23	23	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1
24	24	1	1	1	1	0	1	1	1	1	0	0	1	1	0	1
25	25	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1
26	26	0	1	0	0	0	0	1	0	0	0	1	1	0	1	0
27	27	0	1	1	1	1	1	1	1	1	0	0	1	0	1	1
28	28	1	0	1	0	1	0	0	1	0	0	1	1	0	1	0
29	29	0	1	1	1	0	0	1	1	1	0	1	1	0	1	0
30	30	1	1	1	1	0	1	1	1	0	0	1	0	0	0	1
31	31	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1
32	32	1	1	1	0	0	1	1	1	1	1	0	1	0	0	1
33	33	0	1	1	1	1	1	1	0	1	0	0	0	0	0	1

Appendix 7.3: Posttest data

X1c_posttest_cleaned.csv

	Participan	Group	Started.on	Complete	Time.take	Score	Q..1..1.00	Q..2..1.00	Q..3..1.00	Q..4..1.00	Q..5..1.00	Q..6..1.00	Q..7..1.00	Q..8..1.00	Q..9..1.00	Q..10..1.0(	Q..11..1.0(	Q..12..1.0(	Q..13..1.0(	Q..14..1.0(	Q..15..1.0(	Q..16..1.0(	Q..17..1.0(
1	1	A-Only	########	########	8 mins 7 s	20	1	1	1	1	0	1	1	0	1	0	0	1	0	1	1	1	0
2	2	A-Only	########	########	5 mins 55	27	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	1
3	3	AV	########	########	8 mins 4 s	25	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1
4	4	A-Only	########	########	20 mins 1	22	0	1	1	0	0	1	1	1	1	0	1	1	0	1	1	1	0
5	5	AV	########	########	13 mins 1	16	0	1	0	0	0	1	0	1	0	0	0	1	0	0	1	1	0
6	6	A-Only	########	########	20 mins 4	24	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1
7	7	A-Only	########	########	18 mins 4	19	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1	0
8	8	A-Only	########	########	4 mins 24	29	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0
9	9	A-Only	########	########	15 mins 2	20	0	1	1	1	1	1	0	1	1	1	0	1	0	0	0	1	1
10	10	A-Only	########	########	10 mins 1	19	0	1	0	0	0	1	0	1	1	0	0	1	0	1	1	1	0
11	11	AV	########	########	11 mins 1	23	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	1	0
12	12	A-Only	########	########	16 mins 1	23	0	1	1	1	1	1	0	0	1	0	0	1	0	0	0	1	1
13	13	A-Only	########	########	13 mins 5	19	0	1	1	1	0	1	1	1	1	1	0	1	0	1	1	0	0
14	14	AV	########	########	17 mins 4	14	0	0	0	0	0	1	0	1	1	0	0	1	0	1	1	0	0
15	15	AV	########	########	8 mins 21	23	0	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1
16	16	A-Only	########	########	15 mins 3	14	0	0	0	1	0	1	1	0	1	0	0	1	1	0	0	1	1
17	17	AV	########	########	9 mins 19	30	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	18	AV	########	########	8 mins 44	25	1	1	1	1	1	0	0	1	1	1	1	0	1	0	1	0	1
19	19	A-Only	########	########	17 mins 5	28	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1
20	20	A-Only	########	########	14 mins 5	23	0	1	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1
21	21	A-Only	########	########	7 mins 25	29	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1
22	22	AV	########	########	11 mins 2	21	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1	1	0
23	23	AV	########	########	14 mins 3	23	1	1	1	1	0	1	0	1	1	0	0	1	1	1	1	1	0
24	24	A-Only	########	########	13 mins 4	24	1	1	1	1	0	1	1	1	0	0	1	1	0	1	1	1	0
25	25	A-Only	########	########	24 mins 2	23	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	0
26	26	AV	########	########	15 mins	9	0	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0
27	27	AV	########	########	21 mins 4	21	0	1	1	1	1	1	0	1	0	1	0	1	0	0	1	1	1
28	28	AV	########	########	23 mins 1	17	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	1	0
29	29	A-Only	########	########	18 mins	15	0	1	0	1	1	1	1	0	1	0	0	1	1	0	0	1	0
30	30	AV	########	########	16 mins 5	21	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	0
31	31	A-Only	########	########	23 mins 5	22	1	1	1	0	0	1	1	1	1	0	0	1	0	1	1	1	1
32	32	AV	########	########	8 mins 5 s	22	1	1	1	0	0	1	1	1	0	0	1	1	0	1	1	1	1
33	33	AV	########	########	18 mins 2	24	1	1	1	0	0	1	0	1	1	1	0	1	1	1	1	1	0
34	34	A-Only	########	########	8 mins 54	21	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1
35	35	AV	########	########	18 mins 2	21	0	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1

	Participan	Q..18..1.00	Q..19..1.00	Q..20..1.00	Q..21..1.00	Q..22..1.00	Q..23..1.00	Q..24..1.00	Q..25..1.00	Q..26..1.00	Q..27..1.00	Q..28..1.00	Q..29..1.00	Q..30..1.00	Q..31..1.00	Q..32..1.00
1	1	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1
2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
3	3	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1
4	4	1	1	1	1	0	1	1	1	0	1	0	1	0	1	1
5	5	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1
6	6	0	1	1	1	1	1	1	0	0	1	1	1	0	0	1
7	7	1	1	1	0	1	1	1	1	0	1	0	1	0	0	1
8	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	9	0	1	1	0	1	1	1	0	1	0	1	1	0	1	0
10	10	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1
11	11	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1
12	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
13	13	1	1	1	0	0	1	1	0	1	1	0	1	0	0	1
14	14	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1
15	15	0	1	1	1	1	1	1	0	0	1	0	1	1	1	1
16	16	0	1	1	0	1	0	1	1	0	0	0	0	0	1	0
17	17	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
18	18	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0
19	19	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
20	20	0	1	1	0	1	1	1	1	1	0	1	1	1	1	0
21	21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22	22	0	1	1	1	1	1	1	0	1	0	1	1	0	1	1
23	23	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1
24	24	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1
25	25	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1
26	26	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0
27	27	0	1	1	0	1	1	1	0	1	0	1	1	0	1	1
28	28	1	1	0	0	1	1	0	1	0	1	1	0	1	1	1
29	29	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0
30	30	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1
31	31	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1
32	32	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1
33	33	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1
34	34	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1
35	35	0	1	1	1	1	1	1	0	0	0	1	1	0	1	0

Appendix 7.4: delayed posttest data

X1c_dposttest_cleaned.csv

	Participan	Group	Started.on	Completed	Time.take	Score	Q.1..1	Q.2..1	Q.3..1	Q.4..1	Q.5..1	Q.6..1	Q.7..1	Q.8..1	Q.9..1	Q.10..1	Q.11..1	Q.12..1	Q.13..1	Q.14..1	Q.15..1	Q.16..1	Q.17..1
1	2	A-Only	#######	#######	5 mins	21	22	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	1	0
2	4	A-Only	#######	#######	6 mins	38	28	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1
3	5	AV	#######	#######	47 mins	1!	24	0	1	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1
4	6	A-Only	#######	#######	5 mins	33	27	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1
5	11	AV	#######	#######	4 mins	50	18	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	0	1
6	16	A-Only	#######	#######	7 mins	10	16	0	0	1	0	1	1	0	0	1	0	0	1	1	0	1	1	1
7	19	A-Only	#######	#######	6 mins	9 s	28	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1
8	21	A-Only	#######	#######	14 mins	2!	20	1	1	1	0	0	1	1	1	0	0	0	1	0	1	1	1	0
9	27	AV	#######	#######	7 mins	36	29	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
10	28	AV	#######	#######	6 mins	2 s	18	1	0	1	1	1	1	0	0	1	1	0	1	0	1	1	1	0

	Participan	Q..18..1	Q..19..1	Q..20..1	Q..21..1	Q..22..1	Q..23..1	Q..24..1	Q..25..1	Q..26..1	Q..27..1	Q..28..1	Q..29..1	Q..30..1	Q..31..1	Q..32..1
1	2	1	1	1	1	0	1	1	1	0	1	1	0	0	0	1
2	4	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
3	5	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1
4	6	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0
5	11	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0
6	16	0	1	1	0	1	0	0	1	0	0	0	1	0	1	1
7	19	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
8	21	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1
9	27	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1
10	28	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1

Appendix 7.5: R script to analyse data in appendices 7.1-7.3, pertaining to Appendix 2

X1c_script.r

```
#Updated 20230829 to include delayed post test.

#This is why data1 refers to data2.csv; data1.csv was made before the delayed post test was completed.

data1 <- read.csv('data2.csv')

pretest <- read.csv('pretest.csv')

posttest <- read.csv('posttest.csv')

delayedposttest <- read.csv('delayed_test.csv')

# enables use of the same R packages I used when coding this data

library(groundhog)

#Need to check which version of R is newest after data is captured

groundhog.day <- '2023-11-01'

#Checked and installed

# These libraries and R versions

groundhog.library("tidyverse", groundhog.day, tolerate.R.version='4.3.2')

# Versions likely to change

groundhog.library("rstan", groundhog.day, tolerate.R.version='4.3.2')

groundhog.library("rstanarm", groundhog.day, tolerate.R.version='4.3.2')

groundhog.library("bridgesampling", groundhog.day, tolerate.R.version='4.3.2')

#library("tidyverse")

#Tidyverse not installed using groundhog because it causes conflicts with RMarkdown at least in the current R installation.

x1c_out_table <- read.csv("x1clessons_cleaned.csv")
```

```
pretest <- read.csv("x1cpretest_cleaned.csv")

posttest <- read.csv("x1cposttest_cleaned.csv")

delayedposttest <- read.csv("x1cdposttest_cleaned.csv")

pretest <- filter(pretest, !Participant %in% c(25, 26, 34, 35))

posttest <- filter(posttest, !Participant %in% c(25, 26, 34, 35))

delayedposttest <- filter(posttest, !Participant %in% c(25, 26, 34, 35))

x1c_out_table <- filter(x1c_out_table, !Participant %in% c(25, 26, 34, 35))

#Updated 20230829 for delayed post test

pretest <- pretest %>%  mutate_at(c(6:38), as.numeric)

posttest <- posttest %>%  mutate_at(c(6:38), as.numeric)

delayedposttest <- delayedposttest %>% mutate_at(c(6:38), as.numeric)

#Pretest vowels

pretest <- mutate(pretest, preTrap = (Q..5..1.00 + Q..6..1.00 + Q..8..1.00 + Q..10..1.00 +
Q..18..1.00 + Q..21..1.00 + Q..25..1.00 + Q..32..1.00))

pretest <- mutate(pretest, preStrut = (Q..1..1.00 + Q..4..1.00 + Q..12..1.00 + Q..16..1.00 +
Q..17..1.00 + Q..26..1.00 + Q..27..1.00 + Q..28..1.00))

pretest <- mutate(pretest, preThought = (Q..2..1.00 +  Q..9..1.00 +  Q..11..1.00 +
Q..15..1.00 +  Q..20..1.00 +  Q..23..1.00 +  Q..24..1.00 +  Q..29..1.00))

pretest <- mutate(pretest, preNurse = (Q..3..1.00 +  Q..7..1.00 +  Q..13..1.00 +  Q..14..1.00
+  Q..19..1.00 +  Q..22..1.00 +  Q..30..1.00 +  Q..31..1.00))

#Post test vowels

posttest <- mutate(posttest, postTrap = (Q..5..1.00 + Q..6..1.00 + Q..8..1.00 + Q..10..1.00 +
Q..18..1.00 + Q..21..1.00 + Q..25..1.00 + Q..32..1.00))

posttest <- mutate(posttest, postStrut = (Q..1..1.00 + Q..4..1.00 + Q..12..1.00 + Q..16..1.00
+ Q..17..1.00 + Q..26..1.00 + Q..27..1.00 + Q..28..1.00))

posttest <- mutate(posttest, postThought = (Q..2..1.00 +  Q..9..1.00 +  Q..11..1.00 +
Q..15..1.00 +  Q..20..1.00 +  Q..23..1.00 +  Q..24..1.00 +  Q..29..1.00))

posttest <- mutate(posttest, postNurse = (Q..3..1.00 +  Q..7..1.00 +  Q..13..1.00 +
Q..14..1.00 +  Q..19..1.00 +  Q..22..1.00 +  Q..30..1.00 +  Q..31..1.00))
```

```
#Delayed post test vowels

delayedposttest <- mutate(delayedposttest, delayTrap = (Q..5..1 + Q..6..1 + Q..8..1 +
Q..10..1 + Q..18..1 + Q..21..1 + Q..25..1 + Q..32..1))

delayedposttest <- mutate(delayedposttest, delayStrut = (Q..1..1 + Q..4..1 + Q..12..1 +
Q..16..1 + Q..17..1 + Q..26..1 + Q..27..1 + Q..28..1))

delayedposttest <- mutate(delayedposttest, delayThought = (Q..2..1 +  Q..9..1 +  Q..11..1 +
Q..15..1 +  Q..20..1 +  Q..23..1 +  Q..24..1 +  Q..29..1))

delayedposttest <- mutate(delayedposttest, delayNurse = (Q..3..1 +  Q..7..1 +  Q..13..1 +
Q..14..1 +  Q..19..1 +  Q..22..1 +  Q..30..1 +  Q..31..1))

# calculate gains

x1c_out_table <- x1c_out_table %>% mutate_at(c(3:14), as.numeric)

x1c_table2 <- mutate(x1c_out_table, gains = (Posttest.total - Pretest.total))

x1c_table2a <- tibble(Participant = posttest$Participant,

                      gainTrap = (posttest$postTrap - pretest$preTrap),

                      gainStrut = (posttest$postStrut - pretest$preStrut),

                      gainThought = (posttest$postThought - pretest$preThought),

                      gainNurse = (posttest$postNurse - pretest$preNurse))

x1c_table2 <- full_join(x1c_table2, x1c_table2a, by = "Participant")

# Maybe the line below is unnecessary because I should already have removed the
participants who did not complete the pretest and post test.

#x1c_table2 <- filter(x1c_table2, !is.na(gains))

#Group tibble for summarising

# x1c_table3a <- summarise(x1c_table2, Mean.Gain = mean(gains), Mean.Trap =
mean(gainTrap), mean.Strut =mean(gainStrut), Mean.Nurse = mean(gainNurse),
Mean.Thought = mean(gainThought), .by = Group)

# The above line did not run because .by does not work with rowwise data frames

x1c_table3a <- x1c_table2 %>% group_by(Group) %>% summarise(Mean.Gain =
mean(gains), Mean.Trap = mean(gainTrap), mean.Strut =mean(gainStrut), Mean.Nurse =
mean(gainNurse), Mean.Thought = mean(gainThought))
```

```
x1c_table3a
groundhog.library("BayesFactor", groundhog.day, tolerate.R.version='4.3.2')
#This would also not install under groundhog
#library("BayesFactor")
set.seed(72)
#bayes-t-tests
#overall gains
audiovisual <- filter(x1c_table2, Group == "AV")
audio_only <- filter(x1c_table2, Group == "A-Only")
tchainsAll <- ttestBF(audiovisual$gains, audio_only$gains)
plot(tchainsAll)
t_gainsAll <- t.test(audiovisual$gains, audio_only$gains)
bayesfactorvalueAll <- extractBF(tchainsAll)
meanGainDiffs <- mean(audiovisual$gains) - mean(audio_only$gains)
#TRAP gains
tchainsTrap <- ttestBF(audiovisual$gainTrap, audio_only$gainTrap)
plot(tchainsTrap)
t_gainsTrap <- t.test(audiovisual$gainTrap, audio_only$gainTrap)
bayesfactorvalueTrap <- extractBF(tchainsTrap)
meanTrapGainDiffs <- mean(audiovisual$gainTrap) - mean(audio_only$gainTrap)
#STRUT gains
tchainsStrut <- ttestBF(audiovisual$gainStrut, audio_only$gainStrut)
plot(tchainsStrut)
t_gainsStrut <- t.test(audiovisual$gainStrut, audio_only$gainStrut)
bayesfactorvalueStrut <- extractBF(tchainsStrut)
```

```
meanStrutGainDiffs <- mean(audiovisual$gainStrut) - mean(audio_only$gainStrut)

#THOUGHT gains

tchainsThought <- ttestBF(audiovisual$gainThought, audio_only$gainThought)

plot(tchainsThought)

t_gainsThought <- t.test(audiovisual$gainThought, audio_only$gainThought)

bayesfactorvalueThought <- extractBF(tchainsThought)

meanThoughtGainDiffs <- mean(audiovisual$gainThought) -
mean(audio_only$gainThought)

#NURSE gains

tchainsNurse <- ttestBF(audiovisual$gainNurse, audio_only$gainNurse)

plot(tchainsNurse)

t_gainsNurse <- t.test(audiovisual$gainNurse, audio_only$gainNurse)

bayesfactorvalueNurse <- extractBF(tchainsNurse)

meanNurseGainDiffs <- mean(audiovisual$gainNurse) - mean(audio_only$gainNurse)

#Tabulate BFs for t-tests

GainDescs <- c("Overall gains", "TRAP gains", "STRUT gains", "THOUGHT gains",
"NURSE gains")

allts <- c(t_gainsAll$statistic, t_gainsTrap$statistic, t_gainsStrut$statistic,
t_gainsThought$statistic, t_gainsNurse$statistic)

allbfs <- c(bayesfactorvalueAll$bf, bayesfactorvalueTrap$bf, bayesfactorvalueStrut$bf,
bayesfactorvalueThought$bf, bayesfactorvalueNurse$bf)

gainsTable <- data.frame(GainDescs, allts, allbfs)

gainsTibble <- gainsTable %>% tibble()

gainsTibble <- rename(gainsTibble, Description = GainDescs)

gainsTibble <- rename(gainsTibble, 't Score' = allts)

gainsTibble <- rename(gainsTibble, 'BayesFactor' = allbfs)

gainsTibble
```

```
x1c_table2 <- mutate(x1c_table2, Group = as.factor(Group))
gainsPlot <-  ggplot(x1c_table2, aes(x = Group, y = gains, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  scale_fill_discrete(name = "Group") +
  theme(axis.ticks.x = element_blank()) +
  ggtitle("Overall gains")
trapPlot <-  ggplot(x1c_table2, aes(x = Group, y = gainTrap, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("TRAP gains")
strutPlot <-  ggplot(x1c_table2, aes(x = Group, y = gainStrut, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("STRUT gains")
thoughtPlot <- ggplot(x1c_table2, aes(x = Group, y = gainThought, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("THOUGHT gains")
nursePlot <-  ggplot(x1c_table2, aes(x = Group, y = gainNurse, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
```

```
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("NURSE gains")
# Saves plots all together in a PNG file
groundhog.library("cowplot", groundhog.day ,tolerate.R.version='4.3.2')
#Also problems with cowplot in groundhog! 20230829
#library("cowplot")
plot_grid(gainsPlot, plot_grid(trapPlot, strutPlot), plot_grid(thoughtPlot, nursePlot), ncol = 1)
ggsave(filename="gains.png")
#Fixed trouble with this code in the lengths of the delayed and post tests. 20230829
x1c_table4 <- x1c_table2 %>% mutate(delayedgains = (Delay.Test.Total - Posttest.total))
x1c_table4 <- filter(x1c_table4, !is.na(delayedgains))
posttest2 <- filter(posttest, Participant == delayedposttest$Participant)
x1c_table4a <- tibble(Participant = delayedposttest$Participant,
                      dgainTrap = (delayedposttest$delayTrap - posttest2$postTrap),
                      dgainStrut = (delayedposttest$delayStrut - posttest2$postStrut),
                      dgainThought = (delayedposttest$delayThought - posttest2$postThought),
                      dgainNurse = (delayedposttest$delayNurse - posttest2$postNurse))
x1c_table4 <- full_join(x1c_table4, x1c_table4a, by = "Participant")
#Group tibble for summarising
x1c_table4b <- x1c_table4 %>% group_by(Group) %>% summarise(Mean.Gain = mean(gains), Mean.Delayed.Gain = mean(delayedgains), Mean.Trap.Gain = mean(gainTrap), Mean.Trap.Delay = mean(dgainTrap), Mean.Strut.Gain = mean(gainStrut), Mean.Strut.Delay = mean(dgainStrut), Mean.Nurse.Gain = mean(gainNurse), Mean.Nurse.Delay = mean(dgainNurse), Mean.Thought.Gain = mean(gainThought), Mean.Thought.Delay = mean(dgainThought))
x1c_table4b
#groundhog.library("BayesFactor", groundhog.day, tolerate.R.version='4.3.0')
```

```
#bayes-t-tests

#overall gains

#This is coming back as saying not enough observations - is that possible?

#20230830 Bug found - referring to invalid objects: was referring to
audiovisual$delayedgains

#but needed to use audiovisual_d$delayedgains and also same for audio_only versus
audio_only_d

#20230830 Bug found - was using x1c_table2 to get the groups but I needed to use
x1c_table4

#20230830 Bug found - was using wrong variable names

set.seed(72)

audiovisual_d <- filter(x1c_table4, Group == "AV")

audio_only_d <- filter(x1c_table4, Group == "A-Only")

tchainsAll_d <- ttestBF(audiovisual_d$delayedgains, audio_only_d$delayedgains)

plot(tchainsAll_d)

t_gainsAll_d <- t.test(audiovisual_d$delayedgains, audio_only_d$delayedgains)

bayesfactorvalueAll_d <- extractBF(tchainsAll_d)

meanDelayDiffs <- mean(audiovisual_d$delayedgains) -
mean(audio_only_d$delayedgains)

#TRAP gains

tchainsTrap_d <- ttestBF(audiovisual_d$dgainTrap, audio_only_d$dgainTrap)

plot(tchainsTrap_d)

t_gainsTrap_d <- t.test(audiovisual_d$dgainTrap, audio_only_d$dgainTrap)

bayesfactorvalueTrap_d <- extractBF(tchainsTrap_d)

meanTrapDelayDiff <- mean(audiovisual_d$dgainTrap) - mean(audio_only_d$dgainTrap)

#STRUT gains

tchainsStrut_d <- ttestBF(audiovisual_d$dgainStrut, audio_only_d$dgainStrut)
```

```
plot(tchainsStrut_d)

t_gainsStrut_d <- t.test(audiovisual_d$dgainStrut, audio_only_d$dgainStrut)

bayesfactorvalueStrut_d <- extractBF(tchainsStrut_d)

meanStrutDelayDiffs <- mean(audiovisual_d$dgainStrut) -
mean(audio_only_d$dgainStrut)

#THOUGHT gains

tchainsThought_d <- ttestBF(audiovisual_d$dgainThought, audio_only_d$dgainThought)

plot(tchainsThought_d)

t_gainsThought_d <- t.test(audiovisual_d$dgainThought, audio_only_d$dgainThought)

bayesfactorvalueThought_d <- extractBF(tchainsThought_d)

meanThoughtDelayDiffs <- mean(audiovisual_d$dgainThought) -
mean(audio_only_d$dgainThought)

#NURSE gains

tchainsNurse_d <- ttestBF(audiovisual_d$dgainNurse, audio_only_d$dgainNurse)

plot(tchainsNurse_d)

t_gainsNurse_d <- t.test(audiovisual_d$dgainNurse, audio_only_d$dgainNurse)

bayesfactorvalueNurse_d <- extractBF(tchainsNurse_d)

meanNurseDelayDiffs <- mean(audiovisual_d$dgainNurse) -
mean(audio_only_d$dgainNurse)

#Tabulate BFs for t-tests

DelayDescs <- c("Overall delayed gains", "delayed TRAP gains", "delayed STRUT gains",
"delayed THOUGHT gains", "delayed NURSE gains")

allts_d <- c(t_gainsAll_d$statistic, t_gainsTrap_d$statistic, t_gainsStrut_d$statistic,
t_gainsThought_d$statistic, t_gainsNurse_d$statistic)

allbfs_d <- c(bayesfactorvalueAll_d$bf, bayesfactorvalueTrap_d$bf,
bayesfactorvalueStrut_d$bf, bayesfactorvalueThought_d$bf, bayesfactorvalueNurse_d$bf)

delayTable <- data.frame(DelayDescs, allts_d, allbfs_d)

delayTibble <- delayTable %>% tibble()
```

```
delayTibble <- rename(delayTibble, Description = DelayDescs)

delayTibble <- rename(delayTibble, 't Score' = allts_d)

delayTibble <- rename(delayTibble, 'BayesFactor' = allbfs_d)

delayTibble

x1c_table2 <- x1c_table2 %>% mutate(Complete = Participant %in% posttest2$Participant
== TRUE)

x1c_table5 <- x1c_table2 %>% group_by(Complete) %>% summarise(Mean.Gain =
mean(gains), Mean.Trap = mean(gainTrap), mean.Strut =mean(gainStrut), Mean.Nurse =
mean(gainNurse), Mean.Thought = mean(gainThought), Mean.Lesson =
mean(Lesson.Mean))

x1c_table5

x1c_table4 <- mutate(x1c_table4, Group = as.factor(Group))

delayPlot <-  ggplot(x1c_table4, aes(x = Group, y = delayedgains, fill = Group)) +

  coord_cartesian(ylim = c(-10, 10)) +

  geom_violin() + theme_classic() +

  scale_fill_discrete(name = "Group") +

  theme(axis.ticks.x = element_blank()) +

  ggtitle("Overall delayed gains")

d_trapPlot <-  ggplot(x1c_table4, aes(x = Group, y = dgainTrap, fill = Group)) +

  coord_cartesian(ylim = c(-10, 10)) +

  geom_violin() + theme_classic() +

  theme(legend.position = "None", axis.ticks.x = element_blank()) +

  ggtitle("TRAP delayed gains")

d_strutPlot <-  ggplot(x1c_table4, aes(x = Group, y = dgainStrut, fill = Group)) +

  coord_cartesian(ylim = c(-10, 10)) +

  geom_violin() + theme_classic() +

  theme(legend.position = "None", axis.ticks.x = element_blank()) +
```

```
  ggtitle("STRUT delayed gains")
d_thoughtPlot <- ggplot(x1c_table4, aes(x = Group, y = dgainThought, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("THOUGHT delayed gains")
d_nursePlot <-  ggplot(x1c_table4, aes(x = Group, y = dgainNurse, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("NURSE delayed gains")
# Saves plots all together in a PNG file
groundhog.library("cowplot", groundhog.day ,tolerate.R.version='4.3.2')
plot_grid(delayPlot, plot_grid(d_trapPlot, d_strutPlot), plot_grid(d_thoughtPlot,
d_nursePlot), ncol = 1)
ggsave(filename="delayed.png")
# make linear model(s)
#Assign number of cores, because my home laptop and work desktop have different
processors.
numbercores <- 1
#Gains, condition, means and sds of lesson score: no interaction vs interaction
model1 <- stan_glm(gains ~ Participant + Group + Lesson.Mean + Lesson.SD,
                   data = x1c_table2, family = gaussian, cores = numbercores,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model2 <- stan_glm(gains ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                   data = x1c_table2, family = gaussian,
```

```
                diagnostic_file = file.path(tempdir(), "df.csv"))
#Delayed gains, condition, means and sds of lesson score: no interaction vs interaction
model3 <- stan_glm(delayedgains ~ Participant + Group + Lesson.Mean + Lesson.SD,
                data = x1c_table4, family = gaussian, cores = numbercores,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model4 <- stan_glm(delayedgains ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
#Gains, condition, participant and individual vowel gains: no interaction vs interaction
model5 <- stan_glm(gains ~ Participant + Group + gainTrap + gainStrut + gainThought +
gainNurse,
                data = x1c_table2, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model6 <- stan_glm(gains ~ Participant + Group * (gainTrap + gainStrut + gainThought +
gainNurse),
                data = x1c_table2, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model7 <- stan_glm(gainTrap ~ Participant + Group + Lesson.Mean + Lesson.SD,
                data = x1c_table2, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model8 <- stan_glm(gainTrap ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                data = x1c_table2, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model9<- stan_glm(gainStrut ~ Participant + Group + Lesson.Mean + Lesson.SD,
               data = x1c_table2, family = gaussian,
               diagnostic_file = file.path(tempdir(), "df.csv"))
```

```
model10 <- stan_glm(gainStrut ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                    data = x1c_table2, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model11 <- stan_glm(gainThought ~ Participant + Group + Lesson.Mean  + Lesson.SD,
                    data = x1c_table2, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model12 <- stan_glm(gainThought ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                    data = x1c_table2, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model13 <- stan_glm(gainNurse ~ Participant + Group + Lesson.Mean + Lesson.SD,
                    data = x1c_table2, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model14 <- stan_glm(gainNurse ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                    data = x1c_table2, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
#Delayed gains, condition, participant and individual vowel delayed gains: no interaction vs interaction
model15 <- stan_glm(delayedgains ~ Participant + Group + dgainTrap + dgainStrut + dgainThought + dgainNurse + gainTrap + gainStrut + gainThought + gainNurse,
                    data = x1c_table4, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model16 <- stan_glm(gains ~ Participant + Group * (gainTrap + gainStrut + gainThought + gainNurse + dgainTrap + dgainStrut + dgainThought + dgainNurse + gainTrap + gainStrut + gainThought + gainNurse),
                    data = x1c_table4, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model17 <- stan_glm(dgainTrap ~ Participant + Group + Lesson.Mean + Lesson.SD,
```

```
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model18 <- stan_glm(dgainTrap ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model19<- stan_glm(dgainStrut ~ Participant + Group + Lesson.Mean + Lesson.SD,
               data = x1c_table4, family = gaussian,
               diagnostic_file = file.path(tempdir(), "df.csv"))
model20 <- stan_glm(dgainStrut ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model21 <- stan_glm(dgainThought ~ Participant + Group + Lesson.Mean + Lesson.SD,
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model22 <- stan_glm(dgainThought ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model23 <- stan_glm(dgainNurse ~ Participant + Group + Lesson.Mean + Lesson.SD,
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model24 <- stan_glm(dgainNurse ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                data = x1c_table4, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
#compare models to get a BF
mod1v2 <- bayes_factor(bridge_sampler(model1), bridge_sampler(model2))
```

```
mod3v4 <- bayes_factor(bridge_sampler(model3), bridge_sampler(model4))

mod5v6 <- bayes_factor(bridge_sampler(model5), bridge_sampler(model6))

mod7v8 <- bayes_factor(bridge_sampler(model7), bridge_sampler(model8))

mod9v10 <- bayes_factor(bridge_sampler(model9), bridge_sampler(model10))

mod11v12 <- bayes_factor(bridge_sampler(model11), bridge_sampler(model12))

mod13v14 <- bayes_factor(bridge_sampler(model13), bridge_sampler(model14))

mod15v16 <- bayes_factor(bridge_sampler(model15), bridge_sampler(model16))

mod17v18 <- bayes_factor(bridge_sampler(model17), bridge_sampler(model18))

mod19v20 <- bayes_factor(bridge_sampler(model19), bridge_sampler(model20))

mod21v22 <- bayes_factor(bridge_sampler(model21), bridge_sampler(model22))

mod23v24 <- bayes_factor(bridge_sampler(model23), bridge_sampler(model24))

allmods <- c(mod1v2$bf, mod3v4$bf, mod5v6$bf, mod7v8$bf, mod9v10$bf,
mod11v12$bf, mod13v14$bf, mod15v16$bf, mod17v18$bf, mod19v20$bf, mod21v22$bf,
mod23v24$bf)

# descs <- c("Gains, condition, participant and individual vowel gains: no interaction vs
interaction",

#         "TRAP gains, condition, participant: no interaction vs interaction",

#         "STRUT gains, condition, participant: no interaction vs interaction",

#         "THOUGHT gains, condition, participant: no interaction vs interaction",

#         "NURSE gains, condition, participant: no interaction vs interaction")

descs <- c("Gains, condition, means and sds of lesson score: no interaction vs interaction",

       "Delayed gains, condition, means and sds of lesson score: no interaction vs
interaction",

       "Gains, condition, participant and individual vowel gains: no interaction vs
interaction",

       "TRAP gains, condition, participant and sds of lesson score: no interaction vs
interaction",
```

```
"STRUT gains, condition, participant and sds of lesson score: no interaction vs interaction",

"THOUGHT gains, condition, participant and sds of lesson score: no interaction vs interaction",

"NURSE gains, condition, participant and sds of lesson score: no interaction vs interaction",

"Delayed gains, condition, participant and individual vowel gains: no interaction vs interaction",

"TRAP delayed gains, condition, participant and sds of lesson score: no interaction vs interaction",

"STRUT delayed gains, condition, participant and sds of lesson score: no interaction vs interaction",

"THOUGHT delayed gains, condition, participant and sds of lesson score: no interaction vs interaction",

"NURSE delayed gains, condition, participant and sds of lesson score: no interaction vs interaction")

modelTable <- data.frame(descs, allmods)

modelTibble <- modelTable %>% tibble()

modelTibble
```

Appendix 8: Data and scripts pertaining to Jones (2024d) [Appendix 4]

Appendix 8.1: Data pertaining to Jones (2024d) [Appendix 4]

1Gdata.txt, 2Gdata.txt, 3Gdata.txt., 4Gdata.txt, 5Gdata.txt, 1Ndata.txt, 2Ndata.txt, 3Ndata.txt, 4Ndata.txt, 5Ndata.txt

Lesson 1 GE Group

4

Actually, before watching this video I didn't know what was happening in the post colonial cities and nations. The difficult part for me to understand is this video was too fast to catch . When I listen to one sentence and I'm sure I can catch 65%~80% of all vocabulary, however it's not 100% moreover of course I can't catch the vocabulary I don't know.

On the other hand, the easy part for me is once I understand what the topic is about then I'm able to analyze and guess what will come next.

16

In my opinion, the difficulty of understanding the video is catching the speaker's unique accent and intonation. Regarding some words, I could not even guess the words. In other words, it was impossible for me to distinguish between proper nouns and more general words. In particular, it might be much harder when a word was cut off. Missing some words led to difficulty of understanding the contents of the entire speech. However, when she emphasized or said more slowly, I believe I could grasp what the speaker says. The reason I think is because each word was pronounced clearly and separately. To overcome the obstacles which I feel by listening to the video, it could be said that getting used to listening to people who have unique accents and intonations is one of the effective ways.

8

I could learn about her hometown, Lego city had street harassment or slums and difficulties to recognition of poor people from governance. I was surprised to the statement that they had private buses since public transportation does not work. It was a little bit easy to imagine the situation from her story. However, it was difficult to clearly understand the content of this video and understand the words with difficult pronunciation. It was difficult to distinguish what this word means.

10

I think there are a huge gap between cities are the problem.

We have to cooperate to help each other more.

2

It was challenging to understand the content of this Ted talk due to her strong accent. Moreover, her talking speed was bit faster than my listening capacity.

3

Focusing on the type of city is interesting for me. it gave me new perspective and new thinking of city.

12

Almost it is difficult for me and I can't understand what she said and her English.

I couldn't catch what is main point of this video and what is main topic in this video.

14

Summering is so difficult. This requires me to have a high listening skill. I can listen to the sound of a word. So I can identify what a word sounds. However, when it comes to sentences, there are many components that don't allow me to listen easily. There happen to be linking of sounds and the disappearances of sounds. It make me have a trouble.

Lesson 2 GE Group

4

It seems like in developed country people have less problem about the birth and it makes us think this issue don't arise in other developing country as well. Therefore, I believe this video is good for us to realize what is happening in developing country and lead us to solve this problem together. Furthermore the easy part for me was the figures and proportion they shows and introduces in this video. On the other hand the difficult part was the intonation this woman has makes me confused a lot.

16

Through watching this video, I could understand the circumstances of childbirth in India in more detail. Also, I thought in order to exchange information, taking advantage of mobile devices is a great idea to make medical treatments more accessible. In addition, it is vital to understand how babies' body organizations grow up. When it comes to difficulty hearing this video, the speaker's accent and intonation is so unique. I guess it is affected by her native language. In particular, I can not understand well when she tend to roll her 'r'.

8

I was surprised to the statement that almost women who are pregnant have medical information these days. However, there are the issues such as the access to the medical institutions in India. It is serious problem in the world. The easy point in this listening was the content. For me, it was not too difficult to imagine what she was talking about, although I could not totally and clearly understand the information. The difficulties were understanding pronunciation with Indian accent. I felt challenge to understand what she said the word because I am not used to the Indian sound. I need practice to clearly understand the Indian-English.

10

I thought that it is amazing the number of children becomes fewer after the mobile services are available in India.

Technology can change the world and we have to use it focusing on positive side of it.

2

It was difficult to understand what the woman was saying in the video due to her strong Indian accent.

3

It was new system for me and that was impressive.

12

Her English has Indian accent, so it was hard to listen sometimes.

I was surprised at the situation of pregnant in India and there are more phones than toilets.

14

I tis difficult for me to listen to her English. This is because I am not familiar with her English accent. To get over the problem, I have to know many kinds of accent in English. It is easy a little bit for me to listen to words.

Lesson 3 GE Group

4

We live in an information age however, sometimes the information regarding policy and environment is not open that much while this information is extremely important for people to tackle social problems. In this video, the transparency of information is regarded as a significant element and I agree with the idea as a citizen living in this world.

The easy part for me to understand was the speaker gave the topic clearly and introduced it in more and more detail as the time passed. Thus I'm able to know the information will be dug deeper in the latter part of the video, so I don't have to be confused a lot.On the other hand, the difficult part for me was the vocabulary I faced. The vocabulary related to the environment was unfamiliar for me.

16

I knew that most clothes stores have benefit to provide consumers with their products at reasonable price. On the other hand, they tend to drive to worse the environmental issues. Thus, from my perspective, recognizing which companies destroy the environment, and how large the impacts are is crucial to reduce the burden to the environment. In addition, as for customers, when we buy something, we have to consider not only the prices but also

whether the company is eco-friendly or not, which can encourage companies to deal with the environmental issues, and to produce more eco-friendly goods. When it comes to easy points to understand this video, I think I could get more words compared to previous video because his pronunciation is not so unique. However, he might say some Chinese words, so in the part including like Chinese, it is difficult for me to understand he is speaking unknown English words or Chinese.

8

I found that in China, there was the application called blue map to show the present situation of specific areas through color coding systems. I thought that it was significant innovation or technology which developed the industry.

For me, it was not easy to listen because it was difficult to clearly understand the content in details. However, the pronunciation of this speaker was a little bit easier to understand than the last video.

10

The circumstance of the pollution became more worth than before.

Moreover, the government does not report this, buyers choose the products that they can buy cheaper.

It escalates circumstances more.

We have to know about this more.

2

Compared to previous materials, his English was easier to understand. Also, topic was interesting which made me want to listen him carefully.

3

It was the first time for me to know that there is an application for pollution. Therefore, it was interesting.

12

It is easy to listen to his speech because it was slowly and there is no strange accent.

However, the topic was simply difficult for me to understand.

14

It is difficult to listen to his English because of differences between English pronunciation I am used to listening to and his English accent. However, I can predict what the speaker is talking about by referring to the pictures in the movie.

Lesson 4 GE Group

4

I didn't think deeply about the numbers behind our stories and actions patterns before I watched this video. After watching the video, I'm actually assuming what I'm thinking and how this affects the action I took in my daily life. Occasionally It helps me to convince myself that I can do better next time since I could find out the main cause which makes my action changes. Therefore, I think this video is helpful for our life and I do like this content.

As for the difficulty, if I only looked at one sentence then it was too difficult for me to catch 100 percent of the sentence. I want to be able to figure out at least 90 percent without another sentence.While the easy part was the interesting and fascinating content it possessed. I totally concentrated on this video since the content is quite interesting without feeling it is boring.

16

I believe that not only seeing the numbers but also analyzing by using the stories behind the numbers are useful to control money, so I would like to be conscious of the stories as well in the future. In addition, he mentioned that he could realize the factual problem by considering the stories. I agree with his opinion, because so as to understand the essence of something, more various perspectives rather than one view to see things before my eyes can be necessary. As for a point that I think is easy, when the speaker punctuated words like speaking more strongly and slowly, I could catch each words clearly. On the other hand, he sometimes linked words, so it is hard for me to realize what he is saying.

8

I was surprised to the connection between tracking something, losing money and losing relationships. Also, I needed to think about how much money I spend in daily life. It was easy to listen to his story because of his accent, intonation and his way of talking. However, it was a little bit difficult to understand the overall content of this video because of the lack of knowledge about this experiment.

10

I thought it is effective to know what I am thinking by tracking how I spent my money.

The example of a woman is very interesting and she spent most of her money on clothes to show her power for everyone.

I want to try this next month.

2

It was challenging to catch up with his fast speaking.

3

This story was quite interesting for me

12

His English sometimes has Canadian? accent, so it is sometimes difficult to listen to the words.

However his speaking is slowly and basically easy to listen to.

14

It is a bit easy for me to listen to his English. I guess this is because he speaks British English. I could not distinguish his English words, comparing to usual. I will practice to listening to English words in daily live.

Lesson 5 GE Group

4

People nowadays don’t understand the weight of words therefore, the slandering on the internet has become a social issue. Anonymously sending messages can take away someone’s life and can result in suiside which is one of the social issues.

As for the difficulty, I had to take my time to understand what the speaker said after listening to it 3 times. Sometimes, I first listen to one word and think this word is a totally new word which I didn’t know. However if I listen to this certain word 3 times at some point I’ll realize this word is the word I know before I listen to this video file.

While the easy part for me was the topic itself.

16

I think it is rude to send massages to famous people’s home, even the contents of them are nice to them since this can bother their families as well. However, I believe that her behavior and response to ignore the hateful massages are correct because criticisms tend to interpret whatever the famous people say. Recently, it has been more common to comment and criticize especially for celebrities on the internet as technology has developed. In my opinion, having own ideas is vital, but it is not correct to deprecate those who different opinions from them. Also, some of criticisms can say bad based on only outside of people like appearance, and gender, which should be fixed. Although I could understand the contents and guess unknown words if I listen to the whole video, I can not recognize when I listen to the part of it. Thus, I realized that contexts are very important to me to predict what the speaker says.

8

I was surprised that she had countless hate mails and threads from a lot of people. Also, I was surprised that she tried to meet the person who gave hate comments the most and they talked as common.

It was easy to imagine the situation from her examples or experiences. However, it was difficult to understand her story in details because of the pronunciation that I am not used to. I need to practice to get used to this kind of pronunciation.

10

I thought that there is a still hate crime that we do not know.

Many people suffer the stereotype that we have to be.

By learning experiences, it will effective to change society's thoughts.

2

It was easier to understand/

3

I impressed her life because if my inbox is full of hate things, I can not overcome. She is a strong woman.

12

I think it is one of the biggest problems that hate mail or something like that kind of things today.

It is hard to deal with it, but her dealing is surprising way and brave things.

Her speaking is slowly and it is easy to understand.

14

It was difficult for me to listen to her English. This is because I am not familiar with her English. I noticed that I had to listen to a range varieties of English.

Lesson 1 PV Group

9

The thing that I felt was difficult is that even though I organized well in terms of note-taking, I tend to think the words that I could not understand while listening. As a consequence, I sometimes forget to focus on listening. There are some words that I could not catch although I know the words.

The thing that I felt was easy is that he gave me examples about the workable city so that I could easily understand what he is talking about. Examples helped me to understand the main point.

11

It was hard for me to listen to the words because he spoke so fast. But the words in the video were not too difficult. I managed to understand the content of the video.

15

The question and the video were extremely difficult for me to understand. The first vowel question was especially complicated because the sound was short and I couldn't even comprehend what he was saying. In terms of video, the content was not so difficult but the way he speaks was unusual, therefore, it was arduous for me.

6

It was difficult for me to hear which vowel is it. Also, his words was connected and difficult to hear.

5

I already put my reaction to the previous questions, but I will write it her as well. He talked about the few walkable cities and how they should increase them because the number of car users are increasing. I think it is necessary to rise the number of walking people7s lights which will make them comfortable walking in the street.

He used some words that I could not quite catch and that was a bit difficult.

13

Through this video, I was thinking that whether Tokyo is a walkable city or not. I think some parts of the Tokyo is quite walkable. I actually have to walk on my way to university for about 20 minutes. Although a large number of people say that Tokyo is narrow, there are sufficient space for pedestrians to walk. Meanwhile, there are not enough space to walk in my local area. People might think that country roads are large however, there is no space that people can walk so that it is incredibly difficult and dangerous for pedestrians to walk the road.

The video was quite difficult to understand for me because it includes some unknown vocabulary and the content was ,mentioned in detail. Although I am getting used to

catching the intonation or pronunciation, some parts of the listening were too complicated to understand.

1

At first, I was misunderstanding. I wrote extremely short reaction in the previous section.

I agree that oversized street make traffic worse and invade the road for walking. This is because I think there are a lot of car-based street in Japan in particular my local area. I feel awkward and dangerous when it comes to walking due to driving car. In contrast, street in Paris is made for human use, so I would like to go there in the future. 3

As for feedback, fundamentally it is relatively easy to understand each sentence he said while it is difficult to understand the flow of his speech sometimes. Also, I have hardly distinguish "walk" and "work".

7

I think that walkable city can be good for many things. The walkable city eliminates the need to use a car in everyday life and reduces the environmental problems caused by car emissions. It will also reduce the number of traffic accidents caused by cars. It was easy to understand how the idea of moving house away from the mills was the catalyst, and then sprawl in many places, which is a detriment to the walkable city.

Lesson 2 PV Group

9

The thing that I felt difficult was that there are some difficult words that I could not catch. A lack of vocabulary skills affects the understanding of the video.

The thing that I felt easy was that a speaker gave an actual example such as the percentage of mortality rate, so I could understand the severe current situation of the American healthcare system for pregnant women.

I was very surprised by the fact that the maternal mortality rate in America is higher than in any other country. Especially women of color are suffering from these preventable accidents. The thing that we can do is to raise our awareness to help them through education. The reason is that education is a fundamental tool to realize the solution for this issue and education can help motivate to realize the importance of women's rights.

11

There were some technical words in the video, therefore I missed some parts. The pacing of the story was just right.

15

The vowel test was slightly easy than before because the audio was clear. However, it is still difficult to comprehend and I did not get great score. The womens pronunciation was nice and clear and easy to understand.

6

Her speech was relatively easy to understand, but there are some medical term that I couldn't understand.

5

I agree with her that people should provide more care and support to women who are pregnant and was pregnant. Since they are still facing difficulty of receiving the support, we should tell the government to invest.

The difficulty is that when she said some of technical term and words from medical field.

13

I knew nothing about pregnant women so that it was incredibly new for me to listen to. I strongly believe that we should pay attention to pregnant women more because they are not so strong and we need to check whether they are ok or not. Likewise in the woman's example, if the mother died when she was having birth to the child, the father will be really sad and it is too cruel for them. In Japan, we usually don't hear that women die when they have birth so that I was a little relieved by that. If it is possible, I think hospitals should offer more opportunities for pregnant women to go to hospital for free because it is necessary for them to take measures.

In this video, there were some medical terms about diseases or symptoms so that it was a little hard for me to understand that part. However, except for that part, the content was clear and I could understand what she wanted to say. I think I could catch more vowels so that it actually made me easier to guess the word or vocabulary even though they were the words that I haven't listened.

1

I was surprised at the percentage of maternal mortality in US since this country has one of the latest medical system in the world. However, the fact is serious than I expected due to economic issue and social unconciousness. Finally, I came to know the situation in Japan as well.

I think that this video is relatively easier than precious one. So, I can say most of this video are easy to understand. But one difficulty that I pick up is the number. It sometimes does not make sense immediately.

7

It is inexcusable that more than half of maternal deaths due to complications could be prevented, and that the failure to do so is due to a lack of quality care and neglect of women.

The speaker's English was easy to understand, but there were a few technical terms that I did not know and I could not catch them.

Lesson 3 PV Group

9

The thing that I felt was easy was the topic was about the environment and economy, so it was not very difficult to understand the content. In this talk, I did not have a difficult word but the speed was sometimes fast so that I could not quite catch certain words or phrases. Also, the accent was British so I had a little difficulty listening.

A speaker talked about issues and important things that we have to think about. While I was listening, she gave me clear problems that we currently face and solutions that can help to improve the economy and environment by providing examples and facts. These things helped me to understand better.

Overall, I realized that we need to deal with climate issues and natural risks together. Protecting our environment will lead to economic growth so that social awareness would be a key term to raise our interests about the environmental protection.

11

I could understand content roughly in the video. But I don't get used to listen English accent.

15

Although the vowel test were difficult, I could get some points. Writing the sentences of what speaker said is somewhat challenging because there is some words that are unclear.

6

She talks very fast, so it was hard to hear all the words by playing audio just once.

5

I think her speech was quite influential since she motivated audiences such as "you might be the one of contributor who can solve free environment". those words actually spread and creates society where people are willing to cooperate and try their best efforts on issue.

Her English has pretty strong British accent which sometimes make me to think carefully so it was quite useful.

13

I believe what she said was quite true. What I was impressed the most was that she said we are required to invest to the economy. Now, I am just a university student and I actually don't know what I can do. The thing that I came up is to get a part time job. In the case that I am working, it means that my activity is relating to the economy.

The speaker in the video was a bit fast so that it took me a few times to understand when she was talking about complicated topic such as economy. I guess I still need to listen to a lot more listening materials.

1

From the beginning to the end, I could not understand what she mainly wanted to insist. As far as I remember, she agreed and rejected against what she said previously. This is why I lost her point. However, I could easily understand what she spoke when I listened each sentence. I think that what made me misunderstand is just structure.

7

It was interesting to hear that the way to get people to care about the environment is to show them that environmentally friendly behavior can have a positive impact on our lives, whether they care about the environment or not.

It would have been difficult to understand this speech if I could not understand the phrase "way around", which means "on the contrary".

Lesson 4 PV group

9

The thing that I felt easy was she gave me some examples of financial abuse and her experiences so that I could understand the content. Her speed and pronunciation were easy to listen to.

The thing that I felt was difficult was she gave me a lot of information so that I had to decide the most important part for a summary.

I realized that many people are struggling with financial abuse and domestic violence. Domestic violence might be solved because it is visible but when it comes to financial abuse, it is invisible, so it is very difficult to notice and help them. We must think about how to help them from financial abuse with the collaboration of NGOs that prove the program for it.

11

The content of the video was not difficult for me to understand. vocabulary in the video was not difficult as well.

15

I think a got good scores than before. The women's pronunciation was clear so it was easy for me to copy and write what she said. However, some part was fast to listen.

6

The way she talks was really clear and easy to understand.

Also, I thought acknowledge the financial literacy is really important to live this life because it's complex and there will be less opportunity to gain some money.

5

I think what her father did was understandable since his money belongs to only himself and it can be private issue which he did not want to his wife to look at. Independency and having a secret about the finance is okay from my perspective because people do not know about anything in the future, in fact, they might get divorce.

First she said "do not tell your mother" was catchy and made me wonder what is that mean then she explained at the end which was really easy to understand..

13

I believe that financial information is incredibly important for people so that I usually don't trust people except for my parents or grandparents. In my family, I actually don't use much money so that my parents are taking care of my financial information. Therefore, I think the relationship of the financial dependency only lasts in the case that you are able to trust the person who manages your money.

Compared to previous video, this was quite easy to understand. I believe that it is because the topic was related finance and all the people have to deal with finance almost everyday. As a difficult part, I actually wondered whether I should include the example in the beginning or not.

1

In this video, I have difficulty of listening connected speech. Some pronunciation was changed due to this and I sometimes lost what the contents are. Despite those mentioned difficulty, I could understand the situation of her and victims of financial abuse. As for the reaction against this, I think that to educate how to use money and financial products at school is the most workable way because people can not live without money, but as she mentioned, this problem is caused by some factors. Hence, I feel this is serious problem.

7

I have learned that it is not good to divide up the roles of the family; you take care of the money and I take care of the housework because it leads to financial dependence.

Lesson 5 PV Group

9

The thing that I felt difficult was depending on the topic that I have a family with, a level of understanding would be different in my case. I need to gain knowledge about various topics even I do not like.

The thing that I felt easy was she gave me the experiences and what journalism is doing to promote a real conversation, so I could understand the important parts of this topic. There are some words that I could not hear but I could understand the resort of sentences so that I could gush the meaning of the words.

In my opinion, even though it is very difficult to understand people who have different thinking, behavior that try to understand them is crucial to deal with social disparities. We need to think about what is the true conversation by engaging with people. This will lead to the solution for the polarizing issue.

11

I managed to understand the content of the latter part, but I am not so sure about the content of the first half. There were some words that were not familiar to me that confused me. That was difficult part.

15

The video that I saw was extremely difficult for me. I could not understand what she was talking about to begin with. There were abundant of vocabularly that I could not comprehend. Some of the question of "identifying words" were easily because she was using simple words.

6

It was very hard to understand context of the video because there are a lot of words that I don't know, and also, she speaks very fast, so I couldn't get all words she said.

5

I truly agree with her that real dialogue will be crucial to solve the gap between each side of believer. They ask questions and try to persuade them properly and they have better understanding to each other.

I think her English was quite easy to understand since her English was American accent.

13

I actually do not want to experience this type of conflict. Not only the election but also it actually happens when people support different football teams. Sometimes people forget what they should consider and they start talking about something completely insane. I guess it would be required to think what is important and try not to hurt people who have different opinions.

It was about election and journalism so that the content was quite difficult for me to understand.

1

I agree with what she thought of journalism. It can be a useful tool to connect people throughout the world, but it enables to devide the world as well.

As for the easy and difficult point, to be honest, most part of this video was relatively easy for me except concluding section. I do not know the reason why I could not listen that part. I guess the vocabulary doesn't matter seriously, but probably connected speech.

7

I learned that in a democracy, it is important for those with different opinions to show interest in each other and to talk without prejudice in order to avoid an unusual polarization of our people.

Appendix 8.2: Script for analysis of Appendix 8.1 pertaining to Jones (2024d) [Appendix 4]

20230818x2aqualscript.R

```
setwd("C:/Users/marcj/Documents/CPD/PhD articles/GELT/GELT Qual analysis/drive-download-20230818T072331Z-001")

x1Gdata <- read.delim(file = "1Gdata.txt")

x2Gdata <- read.delim(file = "2Gdata.txt")

x3Gdata <- read.delim("3Gdata.txt")

x4Gdata <- read.delim("4Gdata.txt")

x5Gdata <- read.delim("5Gdata.txt")

x1Ndata <- read.delim("1Ndata.txt")

x2Ndata <- read.delim("2Ndata.txt")

x3Ndata <- read.delim("3Ndata.txt")

x4Ndata <- read.delim("4Ndata.txt")

x5Ndata <- read.delim("5Ndata.txt")

library(tidyverse)

Gbundle <- tibble(c(x1Gdata, x2Gdata, x3Gdata, x4Gdata, x5Gdata))

Nbundle <- tibble(c(x1Ndata, x2Ndata, x3Ndata, x4Ndata, x5Ndata))

library(stringr)

detectandcount <- function(s, p){

 x <- str_detect(s, p)

 if(x == TRUE) {

  return(str_count(s, pattern = p))

 }

 else{
```

```
   return(0)
  }
}
# Gfast <- str_detect(Gbundle, "fast")
# Nfast <- str_detect(Nbundle, "fast")
# Gslow <- str_detect(Gbundle, "slow")
# Nslow <- str_detect(Nbundle, "slow")
# Gfast_count <- str_count(Gbundle, fast)
# Nfast_count <- str_count(Nbundle, fast)
Gfast <- detectandcount(Gbundle, "fast")
Nfast <- detectandcount(Nbundle, "fast")
Gslow <- detectandcount(Gbundle, "slow")
Nslow <- detectandcount(Nbundle, "slow")
Gslow_count <- str_count(Gbundle, slow)
Nslow_count <- str_count(Nbundle, slow)
```

Appendix 9: Data and scripts pertaining to Jones (2024e) [Appendix 5]

Appendix 9.1: Lesson data

Lesson_data_cleaned.csv

	Participan	Group	Pretest_S	Posttest_S	L1_Score	L2_Score	L3_Score	L4_Score	L1	L2	L3	L4	Pre_time	Post_time	Intp_1	Int1_2	Int2_3	Int3_4	Int4_p
1	1	Immediate	14	16	44.44	11.11	22.22	22.22	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
2	1	Immediate	14	16	44.44	11.11	22.22	22.22	NA	NA	NA	NA	NA	#######	NA	NA	NA	NA	NA
3	7	Immediate	27	25	44.44	55.56	44.44	33.33	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
4	10	Immediate	16	25	55.56	16.67	44.44	44.44	#######	#######	NA	NA	#######	#######	13.95625	6.953472	NA	NA	NA
5	10	Immediate	16	25	55.56	16.67	44.44	44.44	#######	#######	NA	NA	NA	#######	NA	6.953472	NA	NA	NA
6	11	Immediate	18	16	0	11.11	-	33.33	#######	#######	NA	#######	#######	#######	6.995139	21.00278	NA	NA	14.03889
7	11	Immediate	18	16	0	11.11	-	33.33	#######	#######	NA	#######	#######	#######	6.995139	7.036111	NA	NA	14.03889
8	13	Immediate	15	19	44.44	33.33	33.33	-	#######	NA	NA	NA	#######	#######	51.52361	NA	NA	NA	NA
9	15	Immediate	8	-	33.33	0	-	22.22	NA	NA	NA	NA	#######	NA	NA	NA	NA	NA	NA
10	16	Immediate	13	11	44.44	-	44.44	33.33	#######	#######	NA	#######	#######	#######	34.5875	0	NA	NA	14.02431
11	16	Immediate	13	11	44.44	-	44.44	33.33	#######	#######	NA	#######	#######	#######	34.5875	0	NA	NA	14.03958
12	17	Immediate	20	21	22.22	33.33	33.33	11.11	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
13	18	Immediate	21	24	55.56	33.33	44.44	44.44	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
14	20	Immediate	17	13	44.44	33.33	33.33	33.33	#######	#######	NA	NA	#######	#######	105.5715	-56.9382	NA	NA	NA
15	20	Immediate	17	13	44.44	33.33	33.33	33.33	#######	#######	NA	NA	#######	#######	105.5715	-56.9389	NA	NA	NA
16	25	Immediate	19	22	33.33	33.33	66.67	55.56	#######	#######	NA	NA	#######	#######	34.86528	0	NA	NA	NA
17	26	Immediate	18	16	11.11	-	44.44	33.33	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
18	27	Immediate	23	22	11.11	-	33.33	44.44	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
19	30	Immediate	24	21	11.11	16.67	-	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
20	31	Immediate	28	22	44.44	33.33	33.33	33.33	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
21	32	Immediate	21	17	33.33	11.11	44.44	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
22	33	Immediate	20	21	22.22	22.22	44.44	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
23	28	Immediate	31	27	-	33.33	-	33.33	#######	NA	NA	NA	#######	#######	6.979861	NA	NA	NA	NA
24	NA	Immediate	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
25	2	Delayed	22	19	22.22	-	33.33	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
26	3	Delayed	20	-	11.11	-	22.22	33.33	NA	NA	NA	NA	#######	NA	NA	NA	NA	NA	NA
27	4	Delayed	18	21	11.11	44.44	33.33	66.67	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
28	5	Delayed	14	16	22.22	-	44.44	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
29	6	Delayed	14	22	22.22	22.22	-	33.33	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
30	8	Delayed	26	23	-	-	22.22	22.22	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
31	8	Delayed	26	23	-	-	22.22	22.22	NA	NA	NA	NA	#######	NA	NA	NA	NA	NA	NA
32	9	Delayed	13	17	33.33	-	-	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
33	9	Delayed	13	17	33.33	-	-	-	NA	NA	NA	NA	NA	#######	NA	NA	NA	NA	NA
34	12	Delayed	19	18	22.22	-	33.33	22.22	#######	#######	#######	#######	#######	#######	20.07292	27.99861	0	0	-0.00069
35	14	Delayed	20	21	11.11	22.22	22.22	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
36	19	Delayed	-	-	-	-	-	-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
37	21	Delayed	16	21	11.11	33.33	33.33	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
38	22	Delayed	-	23	44.44	-	22.22	44.44	#######	#######	#######	#######	NA	#######	NA	28.00486	0.017361	-0.01736	7.015278
39	22	Delayed	-	23	44.44	-	22.22	44.44	#######	#######	#######	#######	NA	#######	NA	28.00417	0.018056	-0.01736	7.015278
40	23	Delayed	18	16	11.11	-	44.44	22.22	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
41	24	Delayed	20	22	22.22	-	33.33	33.33	#######	NA	#######	#######	#######	#######	27.98264	NA	NA	-7.00556	7.011111
42	29	Delayed	11	16	33.33	-	-	11.11	NA	#######	NA	NA	#######	#######	NA	NA	NA	NA	NA
43	34	Delayed	17	18	44.44	44.44	22.22	-	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
44	35	Delayed	21	23	33.33	-	22.22	11.11	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
45	35	Delayed	21	23	33.33	-	22.22	11.11	NA	NA	NA	NA	NA	#######	NA	NA	NA	NA	NA
46	36	Delayed	29	31	22.22	44.44	33.33	44.44	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
47	37	Delayed	24	23	33.33	-	44.44	55.56	NA	NA	NA	NA	#######	#######	NA	NA	NA	NA	NA
48	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Appendix 9.2: Pretest data

X3cpretest_cleaned.csv

	Participar	Started.or	Complete	Time.take	Score	Q..1..1.00	Q..2..1.00	Q..3..1.00	Q..4..1.00	Q..5..1.00	Q..6..1.00	Q..7..1.00	Q..8..1.00	Q..9..1.00	Q..10..1.0(	Q..11..1.0(	Q..12..1.0(	Q..13..1.0(	Q..14..1.0(	Q..15..1.0(	Q..16..1.0(	Q..17..1.0(	Q..18..1.0(
1	1	########	########	9 mins 37	14	1	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1
2	2	########	########	6 days 19	22	0	1	1	1	1	0	0	0	1	0	0	1	0	1	1	1	1	1
3	3	########	########	14 mins 5	20	0	0	1	0	0	1	0	1	0	0	1	1	1	1	1	1	1	1
4	4	########	########	11 mins 6	18	0	0	1	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0
5	5	########	########	8 mins 58	14	0	0	0	1	0	0	0	1	1	0	0	1	0	0	1	1	0	1
6	6	########	########	11 mins 1	14	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1
7	7	########	########	7 mins 47	27	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1
8	8	########	########	5 hours 3(	26	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1
9	9	########	########	9 mins 4	13	0	1	1	0	0	0	0	1	1	0	0	1	0	1	0	0	1	0
10	10	########	########	14 mins 1	16	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	1
11	11	########	########	12 mins 3	18	0	1	1	0	0	0	1	0	1	0	0	0	1	0	1	1	0	1
12	12	########	########	17 hours 4	19	1	1	1	0	0	0	1	1	1	0	0	0	1	0	1	1	0	1
13	13	########	########	7 mins 52	15	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0
14	14	########	########	14 mins 3	20	0	0	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0	1
15	15	########	########	10 hours (	8	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
16	16	########	########	8 mins 45	13	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0
17	17	########	########	1 day	20	1	1	1	0	1	1	1	0	0	0	0	1	0	1	1	1	1	1
18	18	########	########	1 hour 37	21	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1	1	1	0
19	20	########	########	29 mins 4	17	0	1	1	0	1	0	0	1	1	1	0	0	1	0	0	1	1	0
20	21	########	########	14 mins 5	16	1	0	1	0	0	1	0	1	0	0	0	1	0	0	1	1	1	0
21	23	########	########	9 mins 18	18	1	1	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1	1
22	24	########	########	9 mins 28	20	1	0	1	0	0	1	0	1	0	0	1	1	1	1	1	1	1	1
23	25	########	########	13 mins 1	19	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	0
24	26	########	########	11 mins 3	18	0	0	1	0	0	1	1	1	0	0	0	1	0	1	1	1	0	1
25	27	########	########	12 mins 4	23	1	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	1	0
26	28	########	########	17 mins 5	31	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
27	29	########	########	6 mins 45	11	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
28	30	########	########	15 mins 3	24	1	1	1	1	0	1	1	0	1	1	0	1	0	1	1	1	1	1
29	31	########	########	4 hours 2	28	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1	1	1	1
30	32	########	########	9 mins 19	21	0	0	1	1	0	1	0	1	1	0	0	1	0	0	1	1	1	1
31	33	########	########	9 mins 16	20	0	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	1
32	34	########	########	6 mins 31	17	0	1	0	1	0	1	0	0	0	0	0	1	1	1	1	1	0	0
33	35	########	########	10 hours 3	21	0	1	1	0	1	1	0	0	1	0	0	1	1	0	1	1	0	1
34	36	########	########	22 mins 2	29	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
35	37	########	########	14 mins 2	24	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	0

	Participan	Q..19..1.0(	Q..20..1.0(	Q..21..1.0(	Q..22..1.0(	Q..23..1.0(	Q..24..1.0(	Q..25..1.0(	Q..26..1.0(	Q..27..1.0(	Q..28..1.0(	Q..29..1.0(	Q..30..1.0(	Q..31..1.0(	Q..32..1.0(	preTrap	preStrut	preThougl	preNurse
1	1	1	1	0	1	0	1	1	0	1	0	1	0	0	0	3	4	4	3
2	2	0	0	1	1	1	1	1	1	0	1	1	1	1	1	5	6	6	5
3	3	1	1	1	1	1	1	1	0	0	0	1	0	0	1	6	3	6	5
4	4	0	1	1	1	1	1	1	1	1	1	1	0	0	1	4	5	6	3
5	5	0	1	1	1	0	1	1	0	0	0	0	1	0	1	5	3	4	2
6	6	1	1	1	0	1	1	0	0	0	0	1	1	1	1	4	2	5	3
7	7	1	1	0	1	1	1	1	1	1	0	1	1	0	1	6	7	8	6
8	8	1	1	1	1	1	1	1	1	1	0	1	0	0	1	7	6	8	5
9	9	1	1	0	1	0	1	1	0	0	0	1	0	0	0	2	2	5	4
10	10	1	1	1	1	1	0	0	1	0	1	0	0	0	1	3	6	4	3
11	11	1	1	1	0	1	1	1	0	1	0	1	0	0	1	6	3	6	3
12	12	1	1	1	0	1	1	1	0	0	1	0	0	0	1	6	4	5	4
13	13	1	0	0	1	1	1	1	1	0	1	1	1	0	1	5	3	3	4
14	14	1	1	1	1	1	1	0	1	0	1	1	0	0	1	6	5	6	3
15	15	0	1	0	1	0	1	1	0	0	0	1	1	0	0	1	0	4	3
16	16	1	1	1	1	1	1	1	0	0	0	1	0	0	0	3	2	5	3
17	17	1	1	1	0	1	1	1	0	0	0	1	0	0	1	6	4	6	4
18	18	1	1	0	1	1	1	0	1	0	1	1	0	1	1	4	6	7	4
19	20	1	0	1	1	0	1	1	1	0	0	1	0	1	0	5	3	4	5
20	21	1	0	1	0	1	0	1	0	1	1	1	0	0	1	5	6	3	2
21	23	1	1	0	0	1	1	1	0	0	0	1	1	1	1	4	4	6	4
22	24	1	1	0	1	1	0	1	0	1	0	0	0	1	1	5	5	4	6
23	25	1	1	1	1	0	1	1	1	0	1	0	0	1	1	5	4	5	5
24	26	1	1	1	0	1	1	1	0	0	1	1	0	0	1	6	3	5	4
25	27	1	1	0	1	1	1	0	1	1	1	1	1	1	1	3	8	7	5
26	28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	7	8	8
27	29	0	0	0	1	1	1	0	1	0	0	1	1	1	1	1	1	5	4
28	30	1	1	1	1	1	1	0	1	0	0	1	1	0	1	5	6	7	6
29	31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7	7	7	7
30	32	1	1	1	1	1	1	1	1	0	1	1	0	0	1	6	6	6	3
31	33	1	1	0	0	1	1	1	1	0	0	0	0	0	1	6	4	6	4
32	34	1	1	1	0	1	1	0	0	0	1	0	1	1	1	3	4	5	5
33	35	1	1	1	1	1	1	1	1	0	1	1	0	0	1	6	4	7	4
34	36	1	1	1	1	1	1	1	1	1	1	1	1	1	0	6	7	8	8
35	37	1	1	1	1	1	1	1	1	0	1	1	1	0	1	6	7	7	4

Appendix 9.3: Posttest data

X3cposttest_cleaned.csv

	Participar	Started.on	Complete	Time.take	Score	Q..1..1.00	Q..2..1.00	Q..3..1.00	Q..4..1.00	Q..5..1.00	Q..6..1.00	Q..7..1.00	Q..8..1.00	Q..9..1.00	Q..10..1.00	Q..11..1.00	Q..12..1.00	Q..13..1.00	Q..14..1.00	Q..15..1.00	Q..16..1.00	Q..17..1.00	Q..18..1.00
1	1	########	########	4 mins 30	16	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	1	1	0
2	2	########	########	6 mins 1 s	19	0	1	1	0	0	1	0	1	0	0	0	1	0	1	1	1	1	1
3	4	########	########	14 mins 3	21	0	0	1	1	1	1	0	1	1	1	0	0	1	0	1	1	0	1
4	5	########	########	10 mins	16	0	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	0	1
5	6	########	########	10 mins 3	22	0	1	1	0	1	1	1	1	0	0	1	1	0	1	1	1	0	1
6	7	########	########	7 mins 33	25	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1
7	8	########	########	14 mins 1	23	1	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	0	1
8	9	########	########	8 mins 25	17	0	1	1	0	1	1	0	0	1	1	0	1	0	0	0	1	0	0
9	10	########	########	10 mins 1	25	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0
10	11	########	########	20 mins 3	16	0	1	1	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0
11	12	########	########	8 mins 51	18	1	1	1	0	0	1	0	1	0	0	0	1	0	1	1	1	0	1
12	13	########	########	7 mins 52	19	0	0	1	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0
13	14	########	########	12 mins 5	21	0	1	1	0	1	1	0	1	1	0	1	1	1	0	1	0	1	0
14	16	########	########	9 mins 25	11	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
15	17	########	########	5 mins 54	21	1	1	1	1	0	1	0	0	1	0	1	1	1	0	1	1	1	1
16	18	########	########	8 mins 45	24	0	1	1	1	1	1	0	1	1	1	0	1	0	0	1	1	1	0
17	20	########	########	12 mins 4	13	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	1	1	0
18	21	########	########	8 mins 5 s	21	1	1	1	1	1	1	0	1	0	0	0	1	0	1	1	1	1	1
19	22	########	########	10 mins 4	23	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	0	1	1
20	23	########	########	7 mins 3 s	16	0	1	1	0	0	0	0	0	1	1	0	1	0	0	1	1	0	0
21	24	########	########	7 mins 43	22	1	1	1	0	0	1	0	1	0	1	1	1	1	1	1	1	1	0
22	25	########	########	9 mins 55	22	0	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	0	0
23	26	########	########	12 mins 3	16	0	0	1	0	1	1	0	1	0	0	0	1	0	1	1	1	0	1
24	27	########	########	10 mins 5	22	0	1	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1	0
25	28	########	########	17 mins 5	27	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
26	29	########	########	8 mins 37	16	0	0	1	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0
27	30	########	########	7 mins 26	21	0	1	1	0	0	1	1	1	0	0	0	1	1	1	1	1	1	1
28	31	########	########	14 mins 4	22	0	1	1	0	0	1	0	1	1	1	0	1	1	0	1	1	1	1
29	32	########	########	10 mins 4	17	0	0	1	0	1	1	0	1	1	0	0	0	0	0	1	1	1	0
30	33	########	########	9 mins 11	21	1	1	0	1	1	0	1	1	1	0	0	1	1	1	1	1	0	1
31	34	########	########	4 mins 34	18	0	1	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	1
32	35	########	########	5 mins 34	23	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	0
33	36	########	########	8 mins 40	31	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
34	37	########	########	10 mins 2	23	1	1	1	1	1	1	0	1	1	0	0	1	1	0	1	0	1	1

	Participan	Q..19..1.00	Q..20..1.00	Q..21..1.00	Q..22..1.00	Q..23..1.00	Q..24..1.00	Q..25..1.00	Q..26..1.00	Q..27..1.00	Q..28..1.00	Q..29..1.00	Q..30..1.00	Q..31..1.00	Q..32..1.00	postTrap	postStrut	postThou	postNurse
1	1	1	1	0	1	1	1	0	1	0	1	1	0	1	1	0	0	0	4
2	2	1	1	0	0	1	1	1	1	0	0	1	0	1	1	5	4	6	4
3	4	0	1	1	1	1	1	0	1	1	1	0	1	1	1	6	4	5	6
4	5	1	1	1	0	1	0	1	0	0	0	1	0	0	1	7	2	5	2
5	6	1	1	1	0	1	1	1	1	1	0	1	0	0	1	7	4	7	4
6	7	1	1	1	1	1	1	1	1	0	0	1	1	0	1	7	4	7	7
7	8	1	1	1	0	1	1	1	1	0	0	1	0	0	1	7	4	8	4
8	9	1	1	0	1	0	0	1	1	0	1	1	0	1	1	5	4	4	4
9	10	1	1	0	1	0	1	1	1	1	1	1	1	1	1	5	6	6	8
10	11	1	0	1	1	0	0	0	1	0	1	1	0	1	1	2	5	4	5
11	12	1	1	1	0	1	1	1	0	0	0	1	0	0	1	6	3	6	3
12	13	1	1	1	1	1	1	1	1	0	1	1	0	1	0	4	4	4	7
13	14	1	1	1	1	1	1	1	1	0	0	1	0	0	1	6	3	8	4
14	16	1	1	1	0	1	1	0	1	0	1	1	0	0	0	1	3	5	2
15	17	1	1	1	0	1	1	0	0	1	0	1	0	0	1	4	6	8	3
16	18	1	1	1	1	1	1	0	1	0	1	1	1	1	1	6	6	7	5
17	20	1	1	0	0	0	1	0	0	0	1	0	1	1	0	2	3	4	4
18	21	1	1	1	0	1	1	0	0	0	1	1	0	0	1	6	6	6	3
19	22	1	1	1	1	0	1	1	1	0	0	1	0	1	1	8	5	5	5
20	23	1	1	0	1	0	1	0	1	0	1	1	0	1	1	2	4	6	4
21	24	1	1	1	1	1	1	1	0	1	0	0	0	0	1	6	5	6	5
22	25	1	1	0	1	1	1	1	1	1	1	1	0	0	1	5	6	6	5
23	26	1	1	0	0	1	1	1	1	0	0	1	0	0	0	5	3	5	3
24	27	1	1	0	1	1	1	0	1	1	1	1	0	1	1	4	6	7	5
25	28	1	1	1	1	1	1	1	0	1	1	1	0	1	1	7	5	8	7
26	29	1	1	1	1	1	0	0	0	0	1	1	1	0	1	4	4	4	4
27	30	1	1	1	1	1	1	1	0	0	0	1	0	0	1	6	3	6	6
28	31	1	1	1	1	1	1	0	1	0	0	1	0	1	1	6	4	7	5
29	32	1	1	1	1	1	1	0	0	0	1	1	0	0	1	5	3	6	3
30	33	1	1	1	0	1	1	0	0	0	1	1	0	0	1	5	5	7	4
31	34	1	1	1	0	1	0	1	0	1	0	1	1	0	1	7	3	5	3
32	35	1	1	1	1	1	1	1	1	0	1	1	1	1	1	5	5	7	6
33	36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	7	8	8
34	37	1	1	0	1	1	1	1	1	0	1	1	0	0	1	6	6	7	4

Appendix 9.4: Script for analysis of Appendix 9.1-9.3 pertaining to Jones (20243) [Appendix 5]

Analysis.r

```
# Analysis

#Reload library after cleaning

library(groundhog)

groundhog.day <- '2023-12-25'

groundhog.library("knitr", groundhog.day, tolerate.R.version='4.3.1')

groundhog.library("tidyverse", groundhog.day, tolerate.R.version='4.3.1')

groundhog.library("BayesFactor", groundhog.day, tolerate.R.version='4.3.1')

groundhog.library("rstan", groundhog.day, tolerate.R.version='4.3.1')

groundhog.library("rstanarm", groundhog.day, tolerate.R.version='4.3.1')

groundhog.library("bridgesampling", groundhog.day, tolerate.R.version='4.3.1')

groundhog.library("cowplot", groundhog.day ,tolerate.R.version='4.3.1')

groundhog.library("lubridate", groundhog.day ,tolerate.R.version='4.3.1')

pretest <- read.csv('x3cpretest_cleaned.csv')

posttest <- read.csv('x3cposttest_cleaned.csv')

data4 <- read.csv('lesson_data_cleaned.csv')

#Need participants only who did pretest and posttest

data4 <- data4 %>% filter(Participant %in% posttest$Participant) # %>% distinct()

data4 <- data4 %>% filter(Participant %in% pretest$Participant)

# New code lines 20240818

data4 <- data4 %>% filter(Participant %in% pretest$Participant)

data4 <- data4 %>% group_by(Participant) %>% arrange(desc(Pre_time), by_group =
TRUE) %>% slice_head()
```

```
data4 <- data4 %>% ungroup()

#Old code again

pretest <- pretest %>% filter(Participant %in% posttest$Participant)

#calculate standard deviation for intervals between lessons and posttest for each participant: Ok 20220426

data4 <- data4 %>% rowwise() %>% mutate(Int_Deviation = sd(c_across(Intp_1:Int4_p), na.rm = TRUE))

data4a <- data4 %>% mutate_at(c(3:20), as.numeric)

data4a <- data4a %>% mutate(gains = (Posttest_Score - Pretest_Score))

pretest_gran <- pretest %>% select(c(Participant, preTrap, preStrut, preThought, preNurse))

posttest_gran <- posttest %>% select(c(Participant, postTrap, postStrut, postThought, postNurse))

gran <- inner_join(pretest_gran, posttest_gran, by = "Participant")

data4a <- inner_join(data4a, gran, by = "Participant")

data4a <- data4a %>% mutate(gainTrap = (postTrap - preTrap))

data4a <- data4a %>% mutate(gainStrut = (postStrut - preStrut))

data4a <- data4a %>% mutate(gainThought = (postThought - preThought))

data4a <- data4a %>% mutate(gainNurse = (postNurse - preNurse))

#Group tibble for summarising

data4c <- group_by(data4a, Group)

data4d <- summarize(data4c, MeanPre = mean(Pretest_Score), MeanPost = mean(Posttest_Score),

            MeanP_1 = mean(Intp_1, na.rm = TRUE), Mean1_2 = mean(Int1_2,  na.rm = TRUE),

            Mean2_3 = mean(Int2_3,  na.rm = TRUE), Mean3_4 = mean(Int3_4,  na.rm = TRUE),

            Mean4_p = mean(Int4_p,  na.rm = TRUE), MeanIntSD = mean(Int_Deviation, na.rm = TRUE))
```

```
data4d

#Get mean and SD

data4a <- data4a %>% rowwise() %>%  mutate(Lesson.Mean =
mean(c_across(L1_Score:L4_Score), na.rm = TRUE))

data4a <- data4a %>% rowwise() %>%  mutate(Lesson.SD =
sd(c_across(L1_Score:L4_Score), na.rm = TRUE))

data4a

# calculate gains

data2out <- ungroup(data4a) %>% filter(!is.na(gains))

#Group tibble for summarising

data3out <- data2out %>% group_by(Group)

data3out <- summarise(data3out, Mean.Gain = mean(gains), Mean.Trap = mean(gainTrap),
mean.Strut =mean(gainStrut), Mean.Nurse = mean(gainNurse), Mean.Thought =
mean(gainThought))

data3out

set.seed(72)

#bayes-t-tests

#overall gains

immediate <- filter(data2out, Group == "Immediate")

delayed <- filter(data2out, Group == "Delayed")

tchainsAll <- ttestBF(immediate$gains, delayed$gains)

plot(tchainsAll)

t_gainsAll <- t.test(immediate$gains, delayed$gains)

bayesfactorvalueAll <- extractBF(tchainsAll)

meanGainDiffs <- mean(immediate$gains) - mean(delayed$gains)

#TRAP gains

tchainsTrap <- ttestBF(immediate$gainTrap, delayed$gainTrap)
```

```
plot(tchainsTrap)

t_gainsTrap <- t.test(immediate$gainTrap, delayed$gainTrap)

bayesfactorvalueTrap <- extractBF(tchainsTrap)

meanTrapGainDiffs <- mean(immediate$gainTrap) - mean(delayed$gainTrap)

#STRUT gains

tchainsStrut <- ttestBF(immediate$gainStrut, delayed$gainStrut)

plot(tchainsStrut)

t_gainsStrut <- t.test(immediate$gainStrut, delayed$gainStrut)

bayesfactorvalueStrut <- extractBF(tchainsStrut)

meanStrutGainDiffs <- mean(immediate$gainStrut) - mean(delayed$gainStrut)

#THOUGHT gains

tchainsThought <- ttestBF(immediate$gainThought, delayed$gainThought)

plot(tchainsThought)

t_gainsThought <- t.test(immediate$gainThought, delayed$gainThought)

bayesfactorvalueThought <- extractBF(tchainsThought)

meanThoughtGainDiffs <- mean(immediate$gainThought) - mean(delayed$gainThought)

#NURSE gains

tchainsNurse <- ttestBF(immediate$gainNurse, delayed$gainNurse)

plot(tchainsNurse)

t_gainsNurse <- t.test(immediate$gainNurse, delayed$gainNurse)

bayesfactorvalueNurse <- extractBF(tchainsNurse)

meanNurseGainDiffs <- mean(immediate$gainNurse) - mean(delayed$gainNurse)

#Tabulate BFs for t-tests

GainDescs <- c("Overall gains", "TRAP gains", "STRUT gains", "THOUGHT gains",
"NURSE gains")
```

```
allts <- c(t_gainsAll$statistic, t_gainsTrap$statistic, t_gainsStrut$statistic,
t_gainsThought$statistic, t_gainsNurse$statistic)

allbfs <- c(bayesfactorvalueAll$bf, bayesfactorvalueTrap$bf, bayesfactorvalueStrut$bf,
bayesfactorvalueThought$bf, bayesfactorvalueNurse$bf)

gainsTable <- data.frame(GainDescs, allts, allbfs)

gainsTibble <- gainsTable %>% tibble()

gainsTibble <- rename(gainsTibble, Description = GainDescs)

gainsTibble <- rename(gainsTibble, 't Score' = allts)

gainsTibble <- rename(gainsTibble, 'BayesFactor' = allbfs)

gainsTibble

#set up plots

data2out <- mutate(data2out, Group = as.factor(Group))

gainsPlot <-  ggplot(data2out, aes(x = Group, y = gains, fill = Group)) +

  coord_cartesian(ylim = c(-10, 10)) +

  geom_violin() + theme_classic() +

  scale_fill_discrete(name = "Group") +

  theme(legend.position = "None", axis.ticks.x = element_blank()) +

  ggtitle("Overall gains")

trapPlot <-  ggplot(data2out, aes(x = Group, y = gainTrap, fill = Group)) +

  coord_cartesian(ylim = c(-10, 10)) +

  geom_violin() + theme_classic() +

  theme(legend.position = "None", axis.ticks.x = element_blank()) +

  ggtitle("TRAP gains")

strutPlot <-  ggplot(data2out, aes(x = Group, y = gainStrut, fill = Group)) +

  coord_cartesian(ylim = c(-10, 10)) +

  geom_violin() + theme_classic() +
```

```
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("STRUT gains")
thoughtPlot <- ggplot(data2out, aes(x = Group, y = gainThought, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("THOUGHT gains")
nursePlot <-  ggplot(data2out, aes(x = Group, y = gainNurse, fill = Group)) +
  coord_cartesian(ylim = c(-10, 10)) +
  geom_violin() + theme_classic() +
  theme(legend.position = "None", axis.ticks.x = element_blank()) +
  ggtitle("NURSE gains")
# Saves plots all together in a PNG file
plot_grid(gainsPlot, plot_grid(trapPlot, strutPlot), plot_grid(thoughtPlot, nursePlot), ncol =
1)
ggsave(filename="gains.png")
#Assign number of cores, because my home laptop and work desktop have different
processors.
numbercores <- 1
#Gains, condition, means and sds of lesson score: no interaction vs interaction
model1 <- stan_glm(gains ~ Participant + Group + Lesson.Mean + Lesson.SD,
                   data = data2out, family = gaussian, cores = numbercores,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model2 <- stan_glm(gains ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                   data = data2out, family = gaussian, cores = numbercores,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
```

```
#Gains, condition, and sds of lesson score: no interaction vs interaction
model1a <- stan_glm(gains ~ Participant + Group + Lesson.SD,
                    data = data2out, family = gaussian, cores = numbercores,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model2a <- stan_glm(gains ~ Participant + Group *  + Lesson.SD,
                    data = data2out, family = gaussian, cores = numbercores,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
#Gains, condition, sd of lesson intervals: no interaction vs interaction
model3 <- stan_glm(gains ~ Participant + Group + Lesson.Mean + Int_Deviation,
                   data = data2out, family = gaussian, cores = numbercores,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model4 <- stan_glm(gains ~ Participant + Group + Lesson.Mean * Int_Deviation,
                   data = data2out, family = gaussian, cores = numbercores,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
#Gains, condition, participant and individual vowel gains: no interaction vs interaction
model5 <- stan_glm(gains ~ Participant + Group + gainTrap + gainStrut + gainThought +
gainNurse,
                   data = data2out, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model6 <- stan_glm(gains ~ Participant + Group * (gainTrap + gainStrut + gainThought +
gainNurse),
                   data = data2out, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model7 <- stan_glm(gainTrap ~ Participant + Group + Lesson.Mean + Lesson.SD,
                   data = data2out, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
```

```
model8 <- stan_glm(gainTrap ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                   data = data2out, family = gaussian,
                   diagnostic_file = file.path(tempdir(), "df.csv"))
model9<- stan_glm(gainStrut ~ Participant + Group + Lesson.Mean + Lesson.SD,
                  data = data2out, family = gaussian,
                  diagnostic_file = file.path(tempdir(), "df.csv"))
model10 <- stan_glm(gainStrut ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                    data = data2out, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model11 <- stan_glm(gainThought ~ Participant + Group + Lesson.Mean + Lesson.SD,
                    data = data2out, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model12 <- stan_glm(gainThought ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                    data = data2out, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model13 <- stan_glm(gainNurse ~ Participant + Group + Lesson.Mean + Lesson.SD,
                    data = data2out, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model14 <- stan_glm(gainNurse ~ Participant + Group * (Lesson.Mean + Lesson.SD),
                    data = data2out, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model7a <- stan_glm(gainTrap ~ Participant + Group + Int_Deviation,
                    data = data2out, family = gaussian,
                    diagnostic_file = file.path(tempdir(), "df.csv"))
model8a <- stan_glm(gainTrap ~ Participant + Group * Int_Deviation,
```

```
                data = data2out, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model9a <- stan_glm(gainStrut ~ Participant + Group + Int_Deviation,
                data = data2out, family = gaussian,
                diagnostic_file = file.path(tempdir(), "df.csv"))
model10a <- stan_glm(gainStrut ~ Participant + Group * Int_Deviation,
                 data = data2out, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
model11a <- stan_glm(gainThought ~ Participant + Group + Int_Deviation,
                 data = data2out, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
model12a <- stan_glm(gainThought ~ Participant + Group * Int_Deviation,
                 data = data2out, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
model13a <- stan_glm(gainNurse ~ Participant + Group + Int_Deviation,
                 data = data2out, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
model14a <- stan_glm(gainNurse ~ Participant + Group * Int_Deviation,
                 data = data2out, family = gaussian,
                 diagnostic_file = file.path(tempdir(), "df.csv"))
# Add to table
#compare models to get a BF
mod1v2 <- bayes_factor(bridge_sampler(model1), bridge_sampler(model2))
mod1av2a <- bayes_factor(bridge_sampler(model1a), bridge_sampler(model2a))
mod3v4 <- bayes_factor(bridge_sampler(model3), bridge_sampler(model4))
```

```
mod5v6 <- bayes_factor(bridge_sampler(model5), bridge_sampler(model6))

mod7v8 <- bayes_factor(bridge_sampler(model7), bridge_sampler(model8))

mod9v10 <- bayes_factor(bridge_sampler(model9), bridge_sampler(model10))

mod11v12 <- bayes_factor(bridge_sampler(model11), bridge_sampler(model12))

mod13v14 <- bayes_factor(bridge_sampler(model13), bridge_sampler(model14))

mod7av8a <- bayes_factor(bridge_sampler(model7a), bridge_sampler(model8a))

mod9av10a <- bayes_factor(bridge_sampler(model9a), bridge_sampler(model10a))

mod11av12a <- bayes_factor(bridge_sampler(model11a), bridge_sampler(model12a))

mod13av14a <- bayes_factor(bridge_sampler(model13a), bridge_sampler(model14a))

allmods <- c(mod1v2$bf, mod1av2a$bf, mod5v6$bf, mod3v4$bf, mod7v8$bf,
mod9v10$bf, mod11v12$bf, mod13v14$bf, mod7av8a$bf, mod9av10a$bf,
mod11av12a$bf, mod13av14a$bf)

#Add to table

descs <- c("Gains, condition, means and sds of lesson score: no interaction vs interaction",

        "Gains, condition, and sds of lesson score: no interaction vs interaction",

        "Gains, condition, sd of lesson intervals: no interaction vs interaction",

        "Gains, condition, participant and individual vowel gains: no interaction vs
interaction",

        "TRAP gains, condition, participant and lesson mean and standard deviation: no
interaction vs interaction",

        "STRUT gains, condition, participant and lesson mean and standard deviation: no
interaction vs interaction",

        "THOUGHT gains, condition, participant and lesson mean and standard deviation:
no interaction vs interaction",

        "NURSE gains, condition, participant and lesson mean and standard deviation: no
interaction vs interaction",

        "TRAP gains, condition, participant and interval standard deviation: no interaction
vs interaction",
```

```
        "STRUT gains, condition, participant and interval standard deviation: no interaction vs interaction",

        "THOUGHT gains, condition, participant and interval standard deviation: no interaction vs interaction",

        "NURSE gains, condition, participant and interval standard deviation: no interaction vs interaction")

modelTable <- data.frame(descs, allmods)

modelTibble <- modelTable %>% tibble()

modelTibble

plot7a <- plot(model7a, col ="#FFFFFF") + theme_cowplot(10)

plot7a <- plot7a  %>% ggplotGrob()

plot8a <- plot(model8a, col ="#FFFFFF") + theme_cowplot(10)

plot8a <- plot8a  %>% ggplotGrob()

plot9a <- plot(model9a, col ="#FFFFFF") + theme_cowplot(10)

plot9a <- plot9a  %>% ggplotGrob()

plot10a <- plot(model10a, col ="#FFFFFF") + theme_cowplot(10)

plot10a <- plot10a  %>% ggplotGrob()

plot11a <- plot(model11a, col ="#FFFFFF") + theme_cowplot(10)

plot11a <- plot11a  %>% ggplotGrob()

plot12a <- plot(model12a, col ="#FFFFFF") + theme_cowplot(10)

plot12a <- plot12a  %>% ggplotGrob()

plot13a <- plot(model13a, col ="#FFFFFF") + theme_cowplot(10)

plot13a <- plot13a  %>% ggplotGrob()

plot14a <- plot(model14a, col ="#FFFFFF") + theme_cowplot(10)

plot14a <- plot14a  %>% ggplotGrob()

model_list <- c(plot7a, plot8a, plot9a, plot10a, plot11a, plot12a, plot13a, plot14a)

label_listr1 <- c("TRAP:No interaction", "TRAP: Interaction")
```

```
label_listr2 <- c("STRUT: No interaction", "STRUT: Interaction")
label_listr3 <- c("THOUGHT: No interaction", "THOUGHT: Interaction")
label_listr4 <- c("NURSE: No interaction", "NURSE: Interaction")
pg2a <- plot_grid(plot7a, plot8a, labels = label_listr1, label_size = 10, align = "hv")
pg2b <- plot_grid(plot9a, plot10a, labels = label_listr2, label_size = 10, align = "hv")
pg2c <- plot_grid(plot11a, plot12a, labels = label_listr3, label_size = 10, align = "hv")
pg2d <-  plot_grid(plot13a, plot14a, labels = label_listr4, label_size = 10, align = "hv")
pg2 <- plot_grid(pg2a, pg2b, pg2c, pg2d, labels = NULL, ncol = 1, nrow = 4, align = "hv")
ggsave2("glmms.png", pg2)
```