

Komplexe Modellierung: Musikerkennung mit Mathematik

1. Einleitung

Shazam ist eine App für das Smartphone, die unbekannte Songs in Sekundenbruchteilen erkennt und dem Nutzer alles über den Musiktitel mitteilt. Mittlerweile ist Shazam eine der bekanntesten und beliebtesten Apps weltweit – täglich wird die Datenbank bis zu 4 Millionen Mal beansprucht! Vor allem die Schnelligkeit, in denen Songs gefunden werden, und die enorm große Datenbank an vorhandenen Musikstücken zeichnet Shazam besonders aus. Doch was hat das mit Mathe zu tun?

Was Shazam so erfolgreich macht, ist die Idee einen akustischen Fingerabdruck von einem Lied zu generieren. Dieser Fingerabdruck ist, genau wie bei Menschen, für jedes Musikstück einzigartig. Die hohe Effizienz des akustischen Fingerabdruck-Algorithmus sowie ein intelligentes Verfahren beim Durchsuchen der Datenbank haben Shazam zum Erfolg gebracht.

Welche mathematischen Schritte genau hinter dem Fingerabdruck-Algorithmus und der Durchsuchung der Shazam-Datenbank stecken, wird im in einem Schülerworkshop hautnah und interaktiv erfahren. Mit einfachen Matlab-Codes und der Hilfe von ein bis zwei wissenschaftlichen Mitarbeitern entdecken die Schülerinnen und Schüler die mathematischen Grundlagen eines Tons und das mathematische Handwerkszeug für den akustischen Fingerabdruck – die Fourieranalyse. Die Schülerinnen und Schüler werden in einem handlungsorientierten Modul selbst einen akustischen Fingerabdruck einer Beispielaufnahme erstellen und diesen in einer programmierten Datenbank durchsuchen. Dafür ist lediglich Vorwissen über Funktionen erforderlich.

2. Wie funktioniert Shazam?

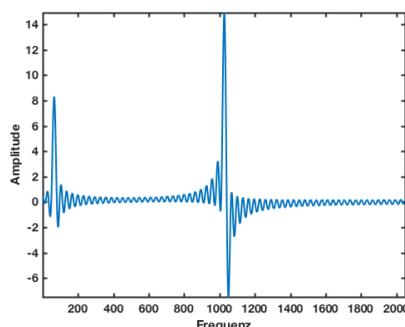


Abbildung 1. Fourier-Transformierte eines Signals mit zwei wesentlichen Frequenzen.

Musikerkennung basiert zunächst auf der Fouriertransformation. Diese zerlegt eine Funktion $f: [0,1] \rightarrow \mathbb{R}$ in ihre Sinus- und Kosinuskomponenten

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(2\pi k t) + b_k \sin(2\pi k t))$$

Für einen kurzen Teil eines Musikstücks gibt sie also an, wieviel Anteil jeweils reine Töne haben (siehe Abbildung 1).

Ein Musikstück kann man nun in kurze Abschnitte unterteilen (dies geschieht durch sog. Hamming-Windows), und in jedem Abschnitt eine Fourier-Analyse durchführen. Daraus erhält man ein Spektrogramm, das zu jedem Zeitfenster die vorhandenen Frequenzen angibt (siehe Abbildung 2 links). Dies ist bereits charakteristisch für eine Aufnahme, also eine Arte Fingerabdruck. Wie für einen tatsächlichen Fingerabdruck werden charakteristische Punkte extrahiert, um diese in einer Datenbank vergleichen zu können. Dies sind im Falle eines Musikstücks die Frequenzen mit der höchsten Amplitude (siehe Abbildung 2 rechts). Dieser Ausschnitt des Liedes muss nun mit der Datenbank abgeglichen werden. Dies geschieht in Shazam durch ein geschicktes Hashing, auf das hier nicht genauer eingegangen wird.

3. Die schnelle Fourier-Transformation

Es ist bereits erkenntlich, dass Shazam wesentlich auf der Fourier-Transformation beruht, welche auf einem Smartphone durchgeführt wird. Die Fourier-Analyse hat viele weitere Anwendungen innerhalb und außerhalb der Mathematik, unter anderem Datenkompression (z.B. arbeiten MP3 und JPEG mit Varianten der Fourier-Transformation). Schneidet man die Fourier-Reihe beim Index n ab, so erfordert die diskrete Fourier-Transformation die Bestimmung von $2n + 1$ Koeffizienten, deren Berechnung jeweils in etwa n Operationen erfordert, mithin insgesamt einen Aufwand von n^2 Operationen. Man sieht sehr schnell dass die Behandlung von Datensätzen mit $n \approx 10^6$ Datenpunkten auf normalen Rechnern unmöglich wird ($1 \text{ GFLOP} = 10^9$ Operationen). Eine solche Datenmenge ist jedoch alles andere als ungewöhnlich. Bei einer Audio-Samplingrate von 48 kHz, d.h. 48.000 Datenpunkten pro Sekunde entsprechen 5 Minuten schon $1,44 * 10^7$ Datenpunkten. Durch geschicktes Rechnen mit komplexen Zahlen (!) lässt sich der Aufwand auf praktisch lineares Verhalten, senken (Fast Fourier Transform, FFT). Erst dies macht Audio- und Videokompression möglich. Die FFT gilt daher zurecht als einer der wichtigsten Algorithmen unserer Zeit. Zugeschrieben wird die Idee Cooley und Tuckey (1965); allerdings gab es die Idee auch schon 1805 von Gauss, um die Bahnen von Asteroiden bestimmen zu können.

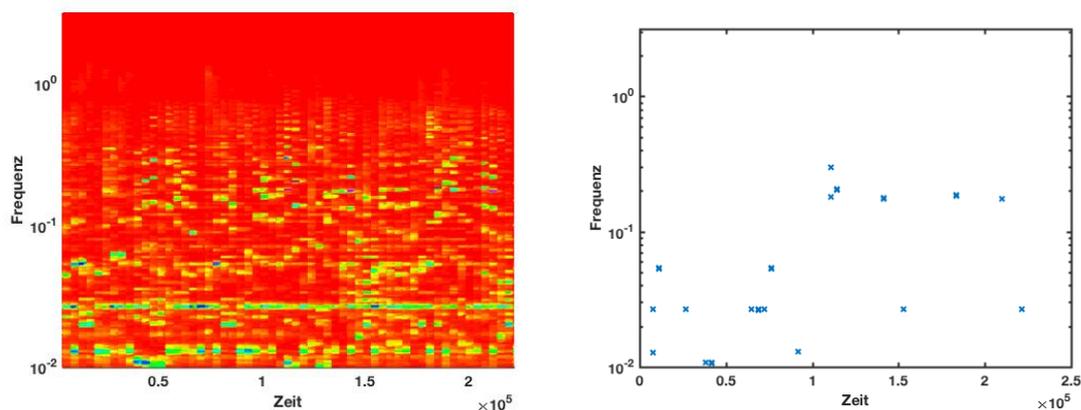


Abbildung 2. Erste Sekunden des Liedes „Pack die Badehose ein“ von Cornelia Froboess: Spektrogramm (links) und daraus extrahierter akustischer Fingerabdruck (rechts).

4. Didaktische Einordnung

Die hier beschriebenen Inhalte werden didaktisch reduziert und aufgearbeitet im Rahmen des Lehr-Lern-Labors CAMMP für mathematische Modellierung der RWTH in einem eintägigen Workshop mit Schülergruppen ab der Jahrgangsstufe acht vermittelt. Es handelt sich um realistische oder angewandte Modellierungsaktivitäten (vgl. Kaiser und Sriraman 2006), denn die zugrundeliegende Problemstellung entstammt der bekannten Anwendung Shazam und ist somit eine reale Forschungsfrage. Aufgrund des hohen Lebensweltbezugs für Schüler/innen erfüllt sie somit die Anforderungen für authentische und relevante Modellierungsaufgaben (vgl. Maaß 2010). Vom Lehr-Lern-Labor CAMMP wurden weitere authentische und relevante Modellierungsaufgaben zum Beispiel zur Geschwindigkeitsmessung, GPS oder Solarenergie entwickelt und für die Behandlung mit Schüler/innen didaktisch aufgearbeitet (vgl. Frank und Roeckerath 2015, Frank, Roeckerath und Hattebuhr 2015).

Literatur

- Frank, M., Roeckerath, C. (2015). Wie kann man mit einer Handykamera Geschwindigkeiten messen? *Der Mathematikunterricht*, 61, 27-31.
- Frank, M., Roeckerath, C., Hattebuhr, M. (2015) Wie funktioniert eigentlich GPS und was hat das mit Mathe zu tun? – Projekttag des EducationLab CAMMP der RWTH-Aachen. *Vortrag, GDM-Tagung*, Basel.
- Frank, M., Roeckerath, C., Hattebuhr, M. (2015) Optimierung der Spiegel in einem Solarkraftwerk – Projekttag des EducationLabs CAMMP der RWTH Aachen. *Vortrag, GDM-Tagung*, Basel.
- Kaiser, G., Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM*, 38(3), 302-310.
- Maaß, K. (2010) Classification scheme for modelling tasks. *Journal für Mathematik-Didaktik*, 31(2): 285-311.

