Spion im Smartphone
Abhören ohne Mikrofon

Lange Nacht der Mathematik 2024

TECHNISCHE UNIVERSITÄT DARMSTADT

Töne messen ohne Mikrofon

In modernen Smartphones lässt sich das Mikrofon meist nicht unbemerkt aktivieren. Dadurch soll das ungewollte Aufnehmen von Audiosignalen verhindert und die Privatsphäre der Nutzenden geschützt werden. Allerdings ist in jedem Smartphone auch ein Beschleunigungssensor verbaut, auf den man fast ohne Restriktionen zugreifen kann. Auch dieser kann zur Tonaufnahme genutzt werden. Erklingt ein Ton, dann vibriert die Luft und der Sensor wird in Schwingung versetzt. Diese Schwingung lässt sich mit dem Beschleunigungssensor messen.

Der Beschleunigungssensor eines Handys misst die Beschleunigung in alle drei Raumrichtungen. Aus der Beschleunigung kann mittels Integration die Amplitude der Schwingung ermittelt werden.

Unser Beschleunigungssensor kann maximal 400 Messungen pro Sekunde ausführen, wir haben also eine Abtastrate von $400\,\mathrm{Hz}$. Gemessen werden können Töne mit einer Frequenz, die maximal halb so groß ist wie die Abtastrate. Das bedeutet, dass der Beschleunigungssensor höhere Töne nicht richtig identifizieren kann.

Fourierreihen: Komplexe Schwingungen zerlegen

Mathematisch kann man Schwingungen vereinfacht durch einen Funktionsgraphen visualisieren, der um den Nullwert oszilliert. Das ist im Grunde genau das Bild, welches unser Beschleunigungssensor sieht! Das heißt: Zu jedem Zeitpunkt können wir durch den Sensor ermitteln, wie sehr das Handy nach oben oder unten ausgelenkt wurde. Nach einer kurzen Zeiteinheit kann die Messung so aussehen wie die schwarze Kurve in Abb. 1.

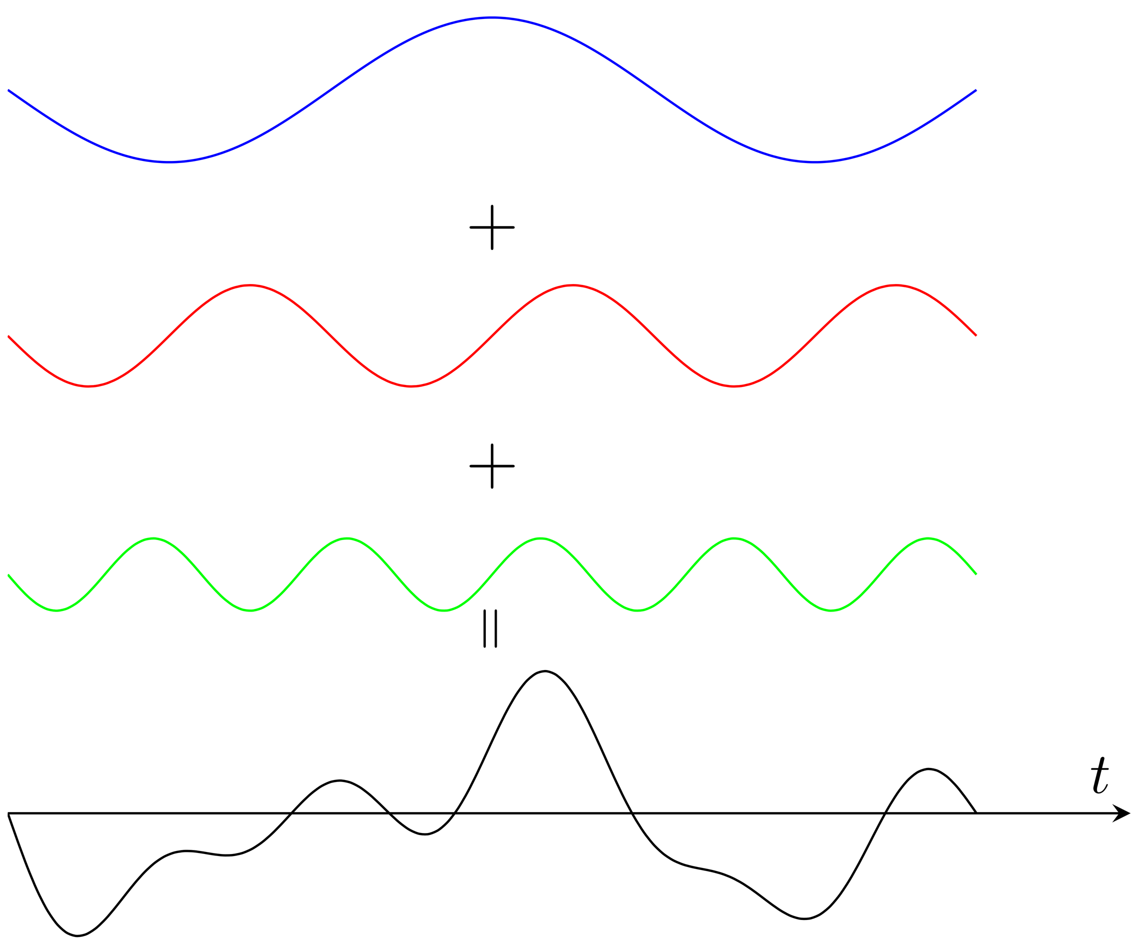

Abb. 1: Durch Vibration verursachte Auslenkung in Schwarz & Zerlegung in Elementarschwingungen (Rot, Blau, Grün)

Das Bild zeigt aber offensichtlich mehr als den eben beschriebenen Funktionsgraphen der Vibration. Wie in der Musik das Lied die Summe aller tönenden Instrumente ist, lassen sich in der Mathematik komplizierte Schwingungen (Abb. 1: schwarzer Graph) zerlegen in eine Summe von (möglicherweise unendlich vielen) Elementarschwingungen (Abb. 1: roter, grüner und blauer Graph). Dabei sagt einem der maximale Ausschlag (Amplitude) einer Elementarschwingung, wie stark sie in der Summe ins Gewicht fällt.

Was wir eben beschrieben haben, ist die sog. Fourierreihendarstellung einer periodischen Funktion f mit Periode T. Formal schreibt man

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(k\omega t) + \sum_{k=1}^{\infty} b_k \sin(k\omega t), \qquad (1)$$

wobei

- $\omega = \frac{2\pi}{T}$ die Grundfrequenz,
- $\cos(k\omega t)$, $\sin(k\omega t)$ die Elementarschwingungen zur Frequenz $k\omega$,
- $2 \cdot a_0$ der Mittelwert der Funktion f und
- $a_1, b_1, a_2, b_2, \ldots$ die Amplituden der Elementarschwingungen

sind.

Amplituden aus Messdaten ermitteln

Tatsächlich sieht der Sensor nicht das kontinuierliche Bild des Funktionsgraphen, welches wir eben skizziert haben, sondern nur eine Annäherung daran. Sensorik funktioniert über Messungen und diese können nur zu endlich vielen Zeitpunkten durchgeführt werden! Für eine große natürliche Zahl N nehmen wir an, dass es $2N$ Sensormessungen gibt. Das heißt, zu den Zeitpunkten $t_1, t_2, \ldots, t_{2N}$ werden die Werte $f(t_1), f(t_2) \ldots, f(t_{2N})$ gemessen und wir setzen $f(t_{2N+1}) = f(t_1)$ für $T = t_{2N+1}$, siehe Abb. 2.

Falls f eine Funktion der Form (1) ist, wollen wir aus den Messdaten den Mittelwert a_0 und die Amplituden $a_1, b_1, a_2, b_2, \ldots$ bestimmen. Mithilfe endlich vieler Daten unendlich viele Unbekannte zu bestimmen ist jedoch unmöglich.

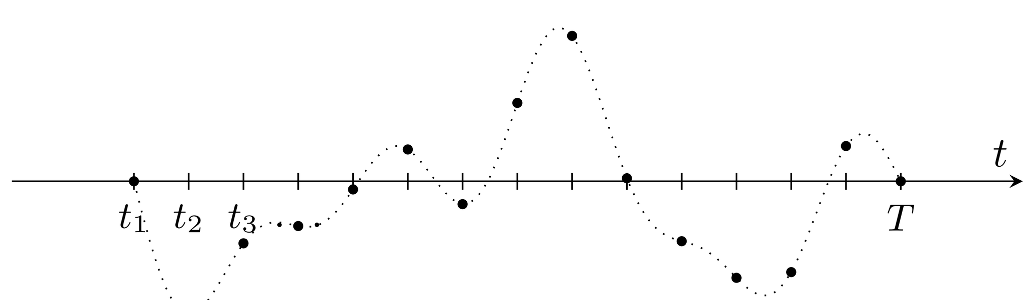

Abb. 2: Durch Vibration verursachte kontinuierliche Auslenkung (gepunktete Kurve) & durch den Sensor gemessene Auslenkungen (Messpunkte) zu den diskreten Zeiten $t_1, t_2, \ldots, T$

Wir behelfen uns mit einer weiteren Annahme und gehen davon aus, dass hochfrequente Elementarschwingungen einen vernachlässigbaren Beitrag zur Fourierreihendarstellung (1) der kontinuierlichen Ausschlagsfunktion f leisten. Da hochfrequente Schwingungen für unsere Anwendung als eine Art Rauschen interpretiert werden können, ist diese Annahme gerechtfertigt.

Insgesamt führt uns das auf ein lineares Gleichungssystem mit $2N$ Gleichungen und $2N$ Unbekannten. Das definiert die sog. diskrete Fouriertransformation:

$$\begin{aligned}
f(t_1) &= a_0 + \sum_{k=1}^{N} a_k \cos(k\omega t_1) + \sum_{k=1}^{N-1} b_k \sin(k\omega t_1) \\
f(t_2) &= a_0 + \sum_{k=1}^{N} a_k \cos(k\omega t_2) + \sum_{k=1}^{N-1} b_k \sin(k\omega t_2) \\
&\vdots \\
f(t_{2N}) &= a_0 + \sum_{k=1}^{N} a_k \cos(k\omega t_{2N}) + \sum_{k=1}^{N-1} b_k \sin(k\omega t_{2N}),
\end{aligned}$$

wobei $2 \cdot a_0 = \frac{1}{2N}(f(t_1) + \ldots + f(t_{2N}))$ dem Mittelwert von f entspricht.

Ein mathematisches Resultat sagt aus, dass die Amplituden mittels

$$a_k = \sum_{n=1}^{2N} f(t_n)\cos(k\omega t_n) \quad \text{und} \quad b_k = \sum_{n=1}^{2N} f(t_n)\sin(k\omega t_n)$$

für $k = 1, \ldots, N$ für a_k und $k = 1, \ldots, N-1$ für b_k berechnet werden können.

Schnelle Fouriertransformation

Möchte man die Koeffizienten $a_1, b_1, a_2, b_2, \ldots, b_{N-1}, a_N$ wie oben beschrieben berechnen, so muss man $(2N)^2$ Multiplikationen durchführen. Für einen Sensor mit Abtastrate $400\,\mathrm{Hz}$ wären das also $400^2 = 160.000$ Multiplikationen.

Mit der sog. schnellen Fouriertransformation (kurz: FFT) geht das deutlich schneller. Sei im Folgenden $2N$ eine Zweierpotenz. Für die FFT wird die Menge der Messpunkte in ungerade und gerade Punkte aufgeteilt

$$\begin{tikzcd}[column sep=tiny]
 & & & \{f(t_1), f(t_2), \ldots, f(t_{2N})\} \arrow[dll] \arrow[drr] & & & \\
 & \{f(t_1), f(t_3), \ldots, f(t_{2N-1})\} \arrow[dl] \arrow[dr] & & & & \{f(t_2), f(t_4), \ldots, f(t_{2N})\} \arrow[dl] \arrow[dr] & \\
\{f(t_1), \ldots, f(t_{2N-3})\} & & \{f(t_3), \ldots, f(t_{2N-1})\} & & \{f(t_2), \ldots, f(t_{2N-2})\} & & \{f(t_4), \ldots, f(t_{2N})\}
\end{tikzcd}$$

Man berechnet nun die schnelle Fouriertransformation der beiden Teilmengen. Dazu teilt man beide Teilmengen wieder in jeweils zwei Teilmengen auf. Dies macht man solange bis man nur noch ein Element pro Teilmenge hat. Um die Fouriertransformation einer größeren Menge zu erhalten, müssen die beiden kleineren Transformationen mit einem Vorfaktor addiert werden.

So müssen insgesamt ungefähr $2N\log_2(2N)$ Multiplikationen durchgeführt werden, anstelle der $4N^2$ Multiplikationen, die für die naive Berechnung nötig sind. Die FFT wird auch in der Funktechnik verwendet, wo große Datenmengen in Echtzeit zerlegt werden müssen (5G arbeitet z.B. mit einer Frequenz von bis zu $43\,\mathrm{GHz}$).

Abtastrate	**Naive Berechnung**	**FFT**
$1\,\mathrm{KHz} = 10^3\,\mathrm{Hz}$	$\mathbf{10^{-6}s}$	$\mathbf{10^{-8}s}$
$1\,\mathrm{MHz} = 10^6\,\mathrm{Hz}$	$1s$	$\mathbf{2 \times 10^{-5}s}$
$1\,\mathrm{GHz} = 10^9\,\mathrm{Hz}$	$10^6 s$	$\mathbf{3 \times 10^{-2}s}$

Abb. 3: Vergleich der Rechenzeiten auf einem typischen Smartphone, in Fett: Echtzeitrechnung möglich

Fachbereich Mathematik der Technischen Universität Darmstadt, AG Numerik und Wissenschaftliches Rechnen

Die lange Nacht der Mathematik

Link zu diesem Poster