Mein A und O

Johannes Barton, Wien 2025

Wenn die Vokale A und O gesprochen bzw. gesungen werden, dann können die Unterschiede nicht nur gehört sondern auch mittels Fourier-Analyse interpretiert werden. In dem vorliegenden Artikel soll eine entsprechende Analyse vorgestellt und die Grundlagen dazu erörtert werden.

Die mathematischen Grundlagen

Fourier–Reihen: Unter recht allgemeinen Bedingungen kann eine periodische Funktion $s(x)$ mit der "Periodenlänge" p in eine Fourier-Reihe mit den Fourier-Koeffizienten c_n entwickelt werden:

$$s(x) = \sum_{n=-\infty}^{+\infty} c_n \exp\left(i2\pi nx/p\right) \quad \text{mit} \quad c_n = \frac{1}{p}\int\limits_{-p/2}^{+p/2} s(x)\exp\left(-i2\pi nx/p\right)\mathrm{d}x$$

Ist $s(x)$ eine reellwertige Funktion, dann gilt offensichtlich $c_n^* = c_{-n}$ für die komplex konjugierten Koeffizienten. Werden die Koeffizienten in die Reihendarstellung eingesetzt, dann ergibt sich

$$s(x) = \frac{1}{p}\int\limits_{-p/2}^{+p/2} s(y)\sum_{n=-\infty}^{+\infty}\exp\left(i2\pi n(x-y)/p\right)\mathrm{d}y$$

und daraus eine Darstellung der δ–Distribution:

$$\delta(x-y) = \frac{1}{p}\sum_{n=-\infty}^{+\infty}\exp\left(i2\pi n(x-y)/p\right)$$

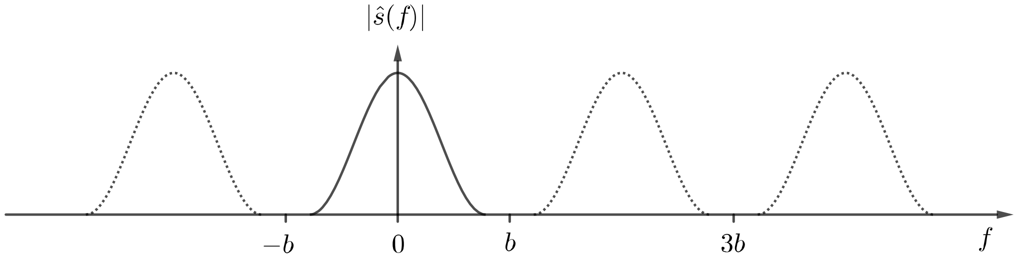
Wird unter x die Zeit t verstanden, dann kann der Ausdruck n/p als Frequenz f angesehen werden. Damit werden alle auftretenden Frequenzen ein ganzzahliges Vielfaches einer Grundfrequenz. Sollen auch aperiodische Funktionen behandelt werden, dann ist ein Grenzübergang $p \to \infty$ durchzuführen. So ergibt sich folgende Fourier-Transformation und deren Inverse:

$$\hat{s}(f) = \int\limits_{-\infty}^{+\infty} s(t)\exp\left(-i2\pi ft\right)\mathrm{d}t \quad \text{und} \quad s(t) = \int\limits_{-\infty}^{+\infty}\hat{s}(f)\exp\left(i2\pi ft\right)\mathrm{d}f$$

Für so eine Fourier-Transformation muss zumindest gefordert werden, dass $|s(t)|$ im Unendlichen verschwindet.

Wenn das Spektrum eines Signals s bandbeschränkt ist, also wenn es eine positive Frequenz b gibt, sodass für $|f| > b$ das Spektrum verschwindet, dann kann das Spektrum periodisch fortgesetzt werden. Dieses periodische Spektrum hätte eine Periodenlänge von $2b$ und kann in eine Fourier-Reihe entwickelt werden.

$$\hat{s}(f) = \frac{1}{2b}\sum_{n=-\infty}^{+\infty} c_n \exp\left(-i2\pi nf/(2b)\right)$$



Es sei angemerkt, dass das Vorzeichen im Exponenten keine Rolle spielt, da sich die Summe über alle ganzen Zahlen erstreckt. Weiters sollte beachtet werden, dass es sich bei $\hat{s}(f)$ um eine Dichte handelt, weil durch die Periodenlänge $2b$ dividiert wurde. Da zum Signal in der Frequenzdarstellung das Signal in der Zeitdarstellung konjugiert ist, liegt die Vermutung nahe, dass es eine exakte Beziehung zwischen den Fourier-Koeffizienten c_n und den Werten des abgetasteten Signals gibt. Dabei soll das Signal in Zeitschritten der konstanten Länge τ abgetastet werden, sodass das abgetastete Signal als Folge der Form $s(n\tau)$ vorliegt. Um die Bedingungen für diese Beziehung zu ermitteln, soll die inverse Fourier-Transformation betrachtet werden:

$$\begin{aligned}
s(t) &= \int\limits_{-\infty}^{+\infty} \frac{1}{2b} \sum_{n=-\infty}^{+\infty} c_n \exp\left(-i2\pi nf/(2b)\right) \exp\left(i2\pi ft\right) \mathrm{d}f \\
&= \sum_{n=-\infty}^{+\infty} c_n \frac{1}{2b} \int\limits_{-b}^{+b} \exp\left(i2\pi f(t - n/(2b))\right) \mathrm{d}f \\
&= \sum_{n=-\infty}^{+\infty} c_n \frac{1}{2b} \frac{2\sin\left(2\pi b(t - n/(2b))\right)}{2\pi(t - n/(2b))} \\
s(t) &= \sum_{n=-\infty}^{+\infty} c_n \frac{\sin\left(\pi(2bt - n)\right)}{\pi(2bt - n)}
\end{aligned}$$

Sei nun k eine ganze Zahl und ein spezieller Zeitpunkt durch $t = k\tau$ gegeben, dann kann die Gleichung

$$s(k\tau) = \sum_{n=-\infty}^{+\infty} c_n \frac{\sin\left(\pi(2bk\tau - n)\right)}{\pi(2bk\tau - n)}$$

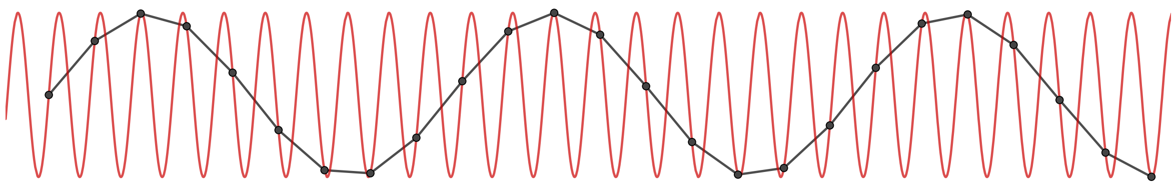
nur für $2b = 1/\tau$ erfüllt werden, da für alle ganzen Zahlen n, k die Beziehung

$$\frac{\sin\left(\pi(k - n)\right)}{\pi(k - n)} = \begin{cases} 1 & \text{falls } k = n \\ 0 & \text{falls } k \neq n \end{cases}$$

gilt. Damit ist schon das *Abtasttheorem* erklärt. Eine exakte Rekonstruktion des Signals aus den Abtastwerten kann durch

$$s(t) = \sum_{n=-\infty}^{+\infty} s(n\tau) \frac{\sin\left(\pi(t - n\tau)/\tau\right)}{\pi(t - n\tau)/\tau}$$

erfolgen, falls die Komponente mit der größten Frequenz mindestens zweimal pro Periode abgetastet wird. Ist dies nicht der Fall, dann treten *Aliasing-Effekte* auf, wie die nachfolgende Darstellung verdeutlicht.



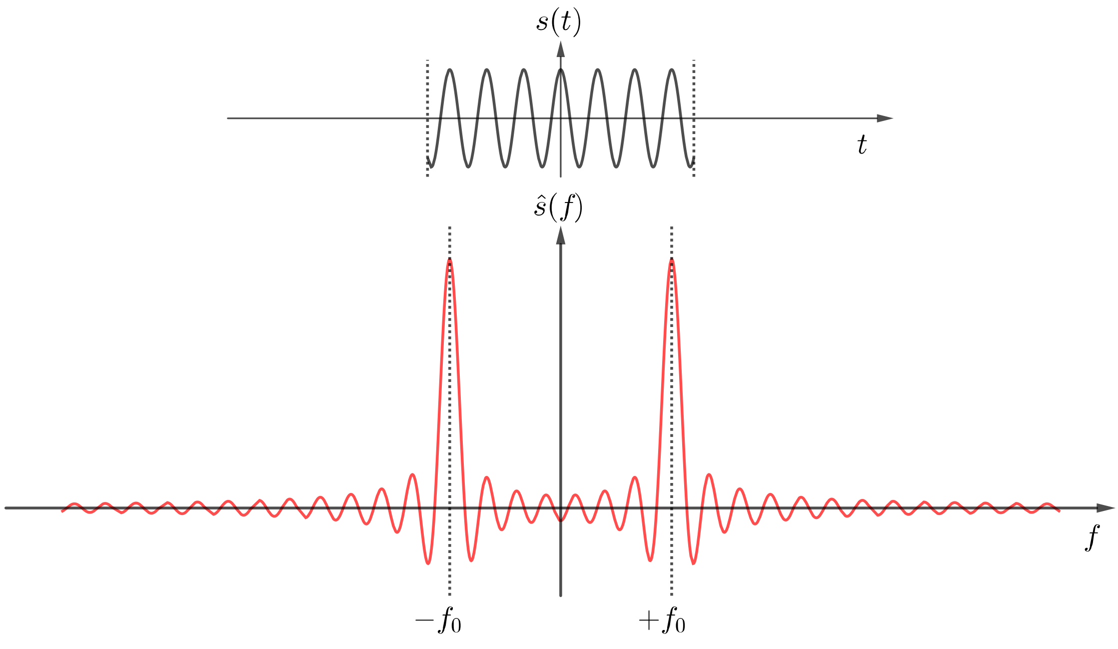
Neben der Problematik des Abtastens sind aber auch Effekte, die aus der Endlichkeit der Messzeit resultieren, nicht zu vernachlässigen. So liefert ein Signal

$$s(t) = \begin{cases} \cos(2\pi f_0 t) & \text{für } |t| \leq T/2 \\ 0 & \text{sonst} \end{cases}$$

die von der Messzeit T abhängige Fourier-Transformierte:

$$\hat{s}(f) = \int_{-T/2}^{+T/2} \cos(2\pi f_0 t) \exp(-i2\pi f t) \mathrm{d}t = \frac{\sin(\pi T (f - f_0))}{2\pi (f - f_0)} + \frac{\sin(\pi T (f + f_0))}{2\pi (f + f_0)}$$

Nachfolgend ist dies für eine Messzeit, die etwas größer als die siebenfache Periodendauer ist, dargestellt.



Nun sind alle Grundlagen gelegt, um die diskrete Fourier–Transformation (DFT) sinnvoll besprechen zu können. Bei dieser Transformation handelt es sich um eine spezielle Abbildung eines N-dimensionalen komplexen Vektorraums auf sich selber.

$$\hat{s}_n = \frac{1}{N} \sum_{k=0}^{N-1} s_k \exp\left(-\frac{i2\pi nk}{N}\right) \quad \text{und} \quad s_k = \sum_{n=0}^{N-1} \hat{s}_n \exp\left(\frac{i2\pi nk}{N}\right)$$

Stellen die s_k das abgetastete Signal in der Zeitdarstellung dar, dann können die $\hat{s}_n$ als Komponenten des Signals in der Frequenzdarstellung gedeutet werden. Auch bei dieser Transformation gilt für reelle Signale, dass $\hat{s}_{N-n}$ die komplex Konjugierte von $\hat{s}_n$ ist.

Damit benötigt das komplexe Spektrum genauso viel Speicherplatz wie das reelle Signal. Da der Kern der DFT gemäß

$$\exp\left(-i2\pi\frac{n+mN}{N}\right) = \exp\left(-i2\pi\frac{n}{N}\right) \quad \text{für alle} \quad m \in \{0, \pm 1, \pm 2, \ldots\}$$

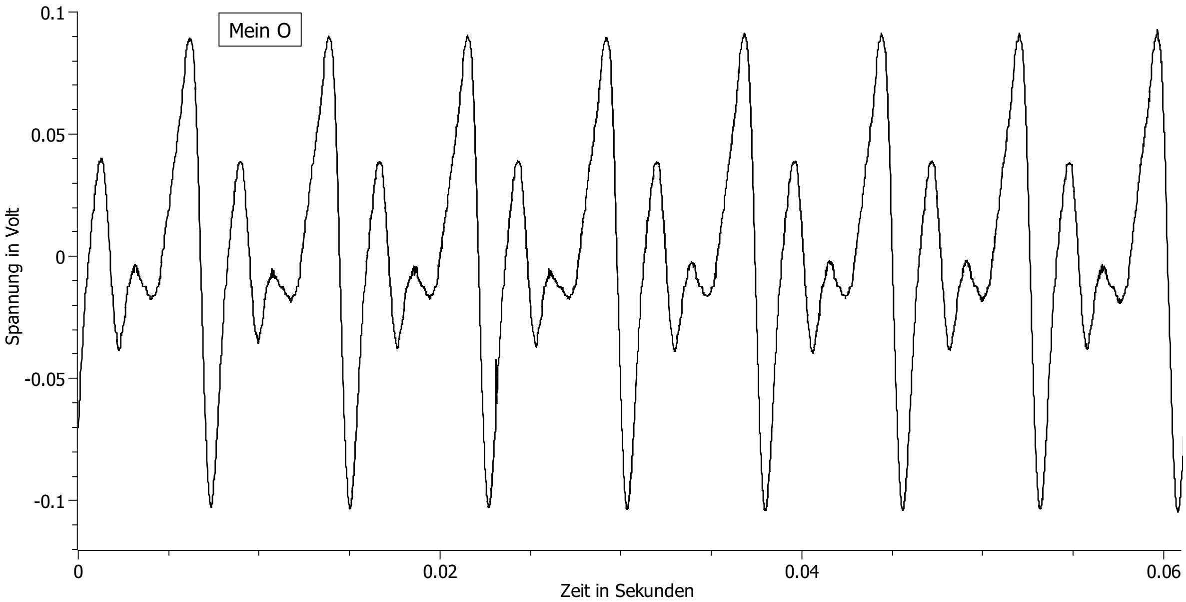
periodisch ist, gilt diese Periodizität sowohl für die Zeit- als auch Frequenzdarstellung. Für die Manipulation von Signalen ist das Faltungsprodukt besonders nützlich. Im diskreten Fall läßt es sich als

$$u_n * v_n = \sum_{m=0}^{N-1} u_m \cdot v_{n-m} = \sum_{m=0}^{N-1} u_{n-m} \cdot v_m$$

anschreiben. Das Faltungstheorem besagt nun, das eine normale Multiplikation in der Zeitdarstellung zu einer Faltung in der Frequenzdarstellung führt und umgekehrt. Soll beispielsweise das Signal in der Zeitdarstellung durch die Faltung mit einer Glättungsmaske geglättet werden, dann entspricht das einer Multiplikation der beiden Spektren.

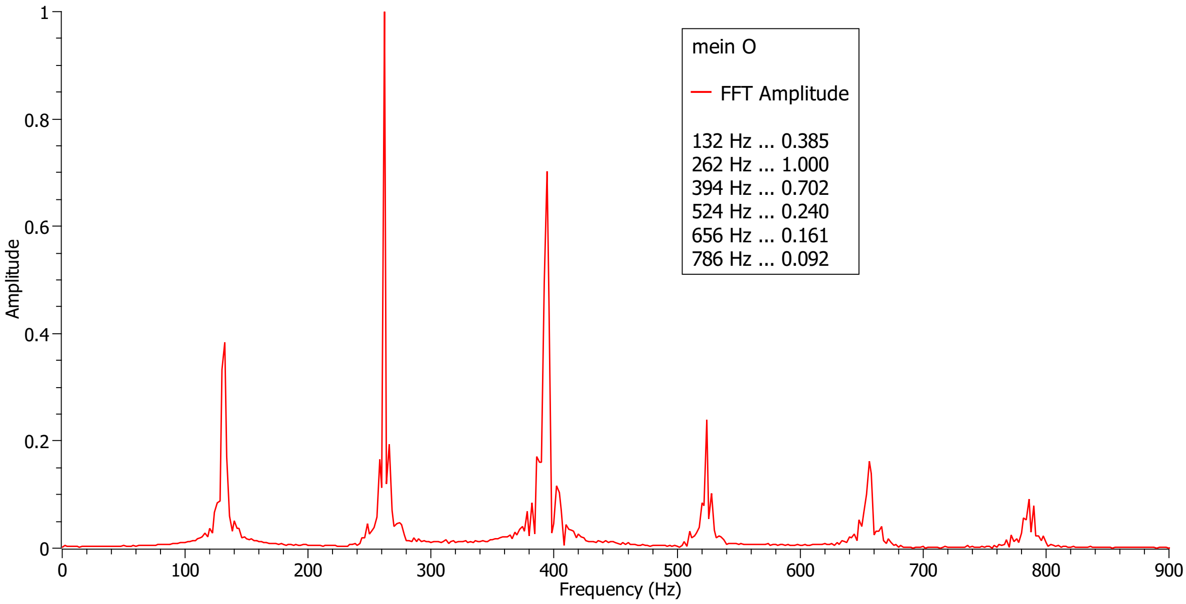
Experiment und Auswertung

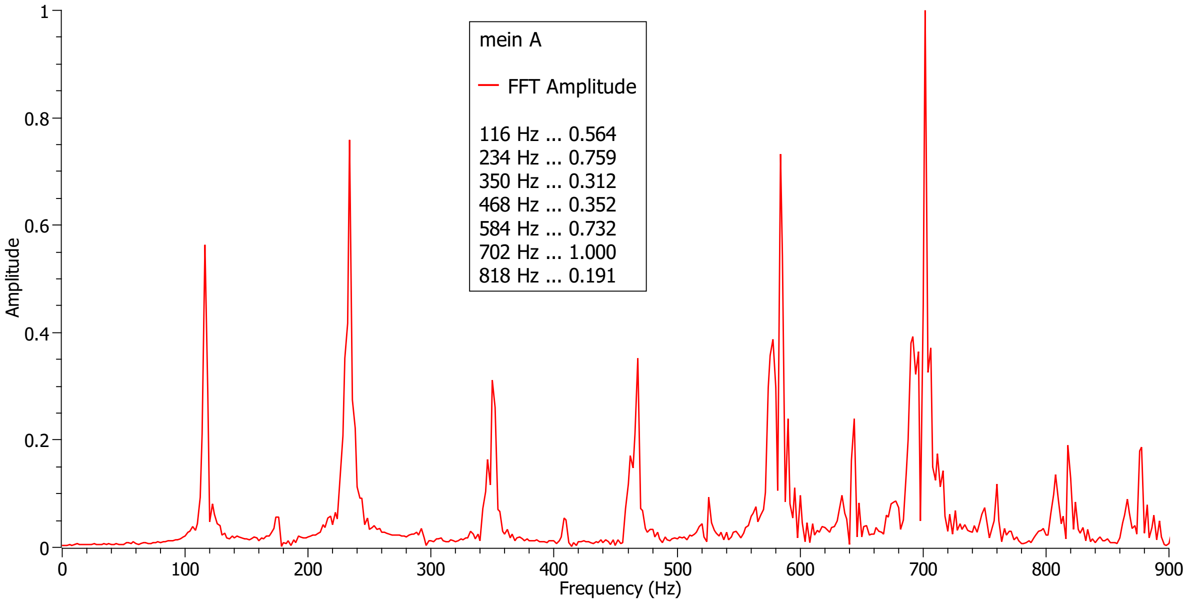
Ein dynamisches Mikrofon mit einer Impedanz von 500 Ω wird direkt an ein digitales Speicheroszilloskop (DSO) angeschlossen. Damit können die gesungenen Vokale A und O aufgezeichnet werden. Bei der verwendeten Zeitbasis arbeitet das DSO mit 1 MHz Sampling Rate und liefert mehr als 500-tausend Daten. Um diese große Datenmenge zu reduzieren wurden je 10 Daten durch Mittelwertbildung zusammengefasst, sodass ein Datenfile mit exakt 50000 Daten zur Auswertung bereit steht. Dadurch wird das Signal mit einer Frequenz von 100 kHz abgetastet. Es kann also davon ausgegangen werden, dass die Bedingungen des Abtasttheorems bei weitem erfüllt sind. Nachfolgend ist exemplarisch ein deutlich verkleinertes Zeitfenster dargestellt.



So ein aus 50000 Daten bestehendes File wurde in `SciDAVis` importiert und eine Fast-Fourier-Transformation (FFT) durchgeführt. Bei einer Abtastfrequenz von 100 kHz wird der Frequenzraum in 2 Hz-Schritten durchlaufen.

Da es in der Akustik nicht auf die Phasenlage der einzelnen Schwingungen ankommt, ist in den nachfolgenden Spektren nur die Amplitude aufgetragen, wobei die größte auf 1 normiert ist. Beide Vokale zeigen deutlich einen Grundton und sogenannte Obertöne. Diese Obertöne sollten ganzzahlige Vielfache vom Grundton sein, was bei einer Frequenzauflösung von 2 Hz durchaus der Fall ist.





Beim Vokal O liegt der Grundton bei (132 ± 1) Hz und der erste Oberton ist der dominanteste.

Das Spektrum des Vokals A ist deutlich kompexer mit einem Grundton von (116 ± 1) Hz und dem fünften Oberton als Dominante. Interessant sind auch die kleineren Peaks, die ziemlich genau zwischen benachbarten Obertönen liegen. Ob in diesem Fall von subharmonischen Tönen gesprochen werden darf, bleibt unbeantwortet.