**Sound waves, Fourier transform, EMD and IMFs**

**Sound waves:**

The vibrations set of from a sound source result in the molecules in the air (or any other transmitting medium) to move and deviate from the original static pressure. The deviations propagate through the air in the form of waves. Therefore, sound is fluctuating air pressure. The pressure in question is very low energy, compared to other forms of energy or even the energy of sea-level air pressure (1/100,000). As waves travel from their source, they weaken and dissipate since no new energy is added to them. Sound waves also reflect off of surfaces in a mirror-like fashion (Heller, 2012; Thewissen & Nummela, 2008).

Sound vibrations just like those of objects, are sinusoidal. At their simplest they are one sine wave while more complex sounds are a combination of sine waves, corresponding to each vibration frequency. The angle of a sinusoid is given by the x axis equalling the frequency times 2 pi. Period of the sinusoid is the time it takes to complete a 360 degree revolution, meaning it equals the inverse of the frequency. Frequency is measured in Hz, meaning cycles per second. A single sinusoid can be denoted as y(t) = A\*cos(2π\*f\*t + φ), where φ is the phase, meaning displacement in time and A is the amplitude (Heller, 2012). Sound travels through air with the same velocity irrespective of wavelength, it therefore, depends on the medium of transmission. Sound wavelength and frequency are inversely related and their product equals sound velocity. For example, sound travels in water five times faster than in air, applying the above relationship, given a single same frequency, the wavelength of a sound inside water will be five times longer than that in air. High and low frequency sounds have short and long wavelengths respectively (Thewissen & Nummela).

Pure, simple tones consist of a single sinusoid, which is perfectly periodic. Most sound waves are complex tones consisting of multiple sinusoids, but this is not to say that complex tones are necessarily aperiodic. If the frequencies of the consisting sinusoids are commensurate, meaning that integer multiples of each frequency are equal to each other (2f1 = 3f2), the complex ton will be periodic. Specifically, the frequency of the period is the greatest common divisor of all the component frequencies. If the greatest common divisor does not exist or if the component frequencies are related through irrational non integral numbers, the tone is aperiodic (Heller, 2012).

**Fourier analysis:**

According to Joseph Fourier any periodic function can be expanded as: y(t) = A1\*cos(2π\*f\*t + φ1) + A2\*cos(2π\*f\*t + φ2) + … + An\*cos(2π\*f\*t + φn). The breaking down of the complex tone starts with the lowest, fundamental frequency and subsequent, higher frequency sinusoids. The transform can also be used on non-periodic waveforms. The Fourier transform describes the amplitude and phase of each sinusoid corresponding to a frequency. Originally proposed in relation to heat flow, its early reception was largely negative. However, it has now become widely used in many fields, including in sound research. The Fourier transform is considered frequency analysis while analysis of hearing based on a waveform is considered a time analysis. A commonly held belief is that the ear automatically performs the Fourier transform on incoming sound waves, a theory suggested by Helmholtz (Heller, 2012; Bracewell, 1989; Malzan, 1989).

According to Helmholtz, the cochlea is a frequency analyser. The different areas of the basilar membrane are tuned to respond to different frequencies; therefore, this would seem to support the idea that the ear performs real time Fourier decomposition. However, analysis of the spiking rates of nerve fibers are not proportional to the Fourier coefficients that correspond to their respective characteristic frequency. That puts into question the idea that the ear conducts a perfect Fourier analysis. Crucially, the cochlea encodes temporal information as well, from the timing of population firing. Cochlear implants electrically stimulate the auditory nerve according to the corresponding auditory signal, meaning they follow the principle of a frequency analysis. Considering the deficits involved in the temporal analysis within the cochlear implant, and the difficulties faced by some cochlear implant users in speech comprehension, it would suggest that indeed the cochlea is not a simple perfect resonator, and modelling implants according to it is sequentially flawed (Rattay, 2000).

Fourier decomposition on a signal results in a sum of sinusoids of different frequencies. In hearing, the inner ear is believed to complete a mechanical form of the Fourier transform by mapping frequencies along the basilar membrane. Alternatively, the Hilbert transform produces a slowly varying envelope and a rapidly varying time structure. Ad neurons that are sensitive to these features have been found in the mammalian ear. Through the Hilbert transform, a signals envelope and fine structure can be inferred using a mathematically rigorous procedure. Discarding time structure information within cochlear implants and using 6 to 8 frequency bands of envelope information does not suffice. Including fine structure information in CI might improve pitch perception and sensitivity. Better pitch perception could in turn lead to better music perception, and speech perception in tonal languages like mandarin Chinese (Smith et al., 2002).

When a complex broadband sound is analyzed in the cochlea of a normal ear, the result is a series of bandpass-filtered signals, each corresponding to one position on the basilar membrane. This aspect of auditory analysis is often modeled (crudely) by short-term Fourier analysis, which expresses the signal in terms of the magnitude and phase of its spectral components. Traditionally, the spectral magnitudes have been regarded as of primary importance for perception, although under some conditions, the phases of the components play an important role (Moore [2002](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR29)). The bandpass signal at a specific place on the basilar membrane (or the signal produced by bandpass filtering to simulate the waveform at one place on the basilar membrane) can be analyzed using the Hilbert transform to create what is called the “analytic signal” (Bracewell [1986](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR2)). The analytic signal can be thought of as a vector that rotates as a function of time; the length of the vector at any time represents the magnitude of the envelope of the signal at that time, and the rate of rotation represents the instantaneous frequency of the signal. In other words, the Hilbert transform can be used to decompose the time signal into its envelope (E; the relatively slow variations in amplitude over time) and temporal fine structure (TFS; the rapid oscillations with rate close to the center frequency of the band). Both E and TFS information are represented in the timing of neural discharges, although TFS information depends on phase locking to individual cycles of the stimulus waveform (Young and Sachs [1979](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR58)). In most mammals, phase locking weakens for frequencies above 4–5 kHz, although some useful phase locking information may persist for frequencies up to at least 10 kHz (Heinz et al. [2001](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR14)). The upper limit of phase locking in humans is not known. Evidence accrued over many years suggests that TFS plays a role in the perception of pitch for both pure and complex tones. For steady complex tones, information from TFS may be important for coding the frequencies of individual resolved partials (Moore et al. [2006b](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR33)) and also for coding the temporal structure of the waveform evoked on the basilar membrane by unresolved harmonics with rank below about 14 (Moore et al. [2006a](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR32); Moore and Moore [2003b](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR40)). For complex tones containing only harmonics above the 14th, the pitch seems to be determined by E rather than by TFS cues (Moore and Moore [2003b](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR40)) and the perceived pitch is relatively weak (Houtsma and Smurzynski [1990](https://link.springer.com/article/10.1007/s10162-008-0143-x#ref-CR17)). Current cochlear implants convey mainly envelope information in different frequency bands and this many partly account for the relatively poor ability of cochlear implanted to understand speech when background sounds are present (Moore, 2008).

**Cochlear implants:**

Cochlear implants have an 80% success in word recognition provided the surrounding environment is ideal, meaning limited environmental noise. They achieve this by electrically stimulating the auditory nerve. The CIS strategy filters a sound into a predetermined number of bands, whose signal is then full-wave rectified and low-pass filtered, to extract the temporal envelope. This is then used to conduct amplitude modulation of a biphasic pulse train that is delivered to a corresponding electrode. Pitch picking is another strategy which uses only selected temporal envelopes from certain bands. A common method to extract the amplitude and frequency modulation is through the Hilbert transform which yields the temporal envelope and the instantaneous frequency. However, the instantaneous frequencies produced often vary rapidly, over a broad range and produce values without physical meaning, therefore, they are difficult to apply to cochlear implants (Nie et al., 2005).

Among the many cochlear implants that have been developed, they all share some common characteristics. They use a microphone to pick up sound, a signal processor to decompose the sound into its frequency components, which are transmitted through a transmission system to implanted electrodes that stimulate the auditory nerve. Depending on the frequency of the signal, electrodes in different positions are stimulated. There are many different types of implants that differ in their transmission links, electrode, and stimulation design and signal processors (Loizou, 1998).

A sound is passed through band-pass filters and envelope detectors or alternatively fast Fourier transform is applied followed by the Hilbert transform. Following the frequency and envelope analysis, a non-linear compressor is used on the envelope whose amplitudes are used to module fixed biphasic carriers. The voltage is converted into pulse trains which are delivered to electrodes arranged in a non-overlapping fashion. This is called Continuous Interleaved Sampling (CIS). Another strategy, similar to the CIS, dynamically selects the bands with the largest envelope amplitudes prior to compression and only corresponding electrodes are stimulated. This strategy is called Advanced Combination Encoder (ACE). Compared to controls, speech perception in cochlear implant users is more challenging (Henry et al., 2023).

Cochlear implant users perform poorly in tasks related to pitch perception, resulting in impaired music perception and enjoyment. The commercially used CIS and ACE strategies employ a division of a signal into a predefined number of frequency bands. Filter bank central frequencies are organized in a linear fashion below 1000 hz and logarithmically above that. The envelope of the filter banks is extracted through half-wave rectification and lowpass filtering and through magnitude response calculation and channel combinations respectively. In CIS all electrodes are stimulated, while in ACE only the channels with maximal magnitudes are simulated. New strategies have been proposed recently, which focus around improving fundamental frequency and music coding (Milczynki et al., 2009).

The cochlear implant’s primary function is to improve speech perception, with less regard towards music perception and enjoyment. Of a total of 35 people who had lost their hearing and received cochlear implants only 16 (46%) continued to listen to music after getting the implant.. Enjoyment of music before losing their hearing was rated on average 8.7 out of 10, while after losing their hearing its was 2.6 out of 10. The enjoyment of music among the 16 who continued listening to it after losing their hearing was on average 5.6 out of 10, which was significantly lower than prior to hearing loss. Interestingly, some gave high cores of enjoyment post implant with one giving a score of 10. It was also found that those who continued listening to music were significantly younger, had higher speech recognition and had been deaf for less time compared to those who didn’t. There was no relationship between listening to music post implant and gender, the elapsed time since getting the implant and the degree of enjoyment prior to loss of hearing. Among the 16 who still listened to music only age was significantly related to enjoyment scores, with speech recognition scores and length of deafness not being significantly related. Of the 35 participants, 69% were disappointed with how music sounded. Of the seven who played music instruments before, only 2 continued playing, of whom one reported disliking the sound of the piano (finding it “awful”) and the other reporting that the glockenspiel had the clearest sound of all his instruments. Of the seven, all listened to musing post implant. None of the participants had professions related to music. A study by Fujita and Ito (1999), suggested that musically trained deaf people are better at music perception post implant due to sue of their previous experience to extract music information from the incomplete information gained from the implant. This is supported by this study too. Participants in this study reported that songs that were known before deafness are best perceived afterwards and visual cues with lyrics aid in understanding the signal. However, a few participants said that music sounded like noise (Mirza et al., 2003).

**Empirical mode decomposition:**

Traditionally, the Fourier analysis has been largely used for energy-frequency distributions so much so that spectrum analysis and the Fourier transform have become almost synonymous. This is due to how effective it is and equally how simple. However, the Fourier transform can only be used when data is linear and either periodic or stationary. Regarding stationarity, many definitions of it exist, some more lenient than others. It can be generally thought of as a variable not changing within a limited time span. Stationarity cannot be established perfectly since data exist within finite time spans, therefore approximation is necessary. Practically in most cases, data is transient in nature. Regarding linearity, linear approximations are effective, however, natural phenomena tend to be non-linear when their variations are finite in amplitude. Despite the non-linearity and stationarity of data, the Fourier transform continues to be used, through the nonchalant acceptance of linearity and stationarity as properties of the processed data.

Were the Fourier transform to be used on non-stationary and non-uniform data, additional harmonic components would be added. However, the addition would be made with the energy being spread over a wide frequency range. This means that the manufactured extra harmonics and the wide frequency spectrum doesn’t not accurately represent the energy density of the signal. The Fourier decomposition, can also require negative intensity signals that cancel out other components, which doesn’t not make sense when considered physically. Furthermore, additional harmonics are necessary to represent deformed wave-profiles which are a result of non-linearities. This means that both non-linearity and non-stationarity can cause the Fourier transform to manufacture additional harmonic components which spread the energy and produce a misleading representation of energy-frequency distributions.

The empirical mode decomposition (EMD) method generates intrinsic mode functions (IMF). The energy associated with intrinsic time scales is extracted and expressed in IMFs whose instantaneous frequencies can be calculated through their Hilbert transforms. The local energy and instantaneous frequencies provide a full energy-frequency-time distribution of the data.

Non-stationary data processing methods are limited to linear systems since they depend on Fourier analysis. Nonetheless they include the spectrogram, which is a limited time window-width Fourier analysis, the wavelet analysis, which is an adjustable window Fourier analysis. The Wigner-Ville distribution is the Fourier transform of the central covariance function, while the evolutionary spectrum is an expanded version of the Fourier transform to include a family of orthogonal functions rather than the sine or cosine. The empirical orthogonal function expansion is a departure from previous Fourier-like methods. All of these methods, however, fall short regarding non-stationarity and as mentioned above are majority linear expansions. For non-stationary data the Fourier transform is physically meaningless.

The notion of instantaneous frequency has been controversial due to the need of an entire oscillation of a sine or cosine wave for the local frequency to be defined, which wouldn’t make sense for non-stationary data whose frequency changes. Instantaneous frequencies also have a set of limitations that they need to follow to decompose data in a meaningful way. The conditions are that the functions are symmetric with respect to the local zero mean, meaning that at any point the mean value of the envelope as defined by the local maxima and local minima is zero, and that they have the same numbers of zero crossings and extrema. Intrinsic mode functions satisfy both conditions. The first requirement is necessary for the instantaneous frequency to be free of fluctuation that have been caused by asymmetric waveforms. For non-stationary data, the local mean is impossible to define, therefore, the local mean of the envelopes defined by the local maxima and minima is used as a substitute to induce local symmetry. The IMF defines harmonic distortions as intrawave frequency modulations which are frequency changes within a wave and are much more physical compared to distortions.

The EMD method is intuitive, direct, a posteriori and adaptive. The decomposition functions with the assumptions that there is at least one maximum and minimum (instead of using two zero-crossings, in case a signal has none), the characteristic time scale is defined by the time lapse between the extrema, and if there are no extrema but there are inflection points, differentiation can be applied for the extrema to be revealed.

The method identifies intrinsic oscillatory modes from their characteristic time scales empirically, and then decomposes the data. The first component is the difference between the data and the mean of the upper and lower envelopes as defined by the local maxima and minima respectively. For the second component, the first component is used in place of the data from which the new local mean is subtracted. The extracted IMFs are then used to compute the instantaneous frequency through the Hilbert transform. The IMF represents a generalized Fourier expansion. The IMF also allows for the amplitude and frequency modulations to be separated, distancing the EMD from the constant amplitude and fixed frequency components of the Fourier expansion (Huang et al., 1998).

References

Heller, E. J. (2012). *Why you hear what you hear: An experiential approach to sound music, and psychoacoustics.* Princeton University Press.

Thewissen , J. G. M., & Nummela, S. (2008). *Sensory evolution of the threshold: Adaptations in secondarily aquatic vertebrates.* University of California Press.

Bracewell, R. N. (1989). The Fourier transform. *Scientific American, 206*(6), 89-95.

Malzan, J. (1989). Analysis of frequency by the ear. *South African Journal of Science, 85*(1), 33.

Rattay, F. (2000). Basics of hearing theory and noise in cochlear implants. *Chaos, Solitons and Fractals, 11*, 1875-1884.

Loizou, F. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*. (The same article can be found with the title: Introduction to cochlear implants, published one year later)

Henry, F., Glavin, M., & Jones, E. (2023). Noise reduction in cochlear implant signal processing: A review and recent developments. *IEEE Reviews in Biomedical Engineering, 16*.

Nie, K., Stickney, G., & Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Transactions in Biomedical Engineering, 52*(1).

Milczynski, M., Wouters, J., & van Wieringen, A. (2009). Improved fundamental frequency coding in cochlear implant signal processing. *The Journal of the Acoustical Society of America, 125*, 2260.

Mirza, S., Douglas, S. A., Lindsey, P., Hildreth, T., & Hawthorne, M. (2003). Appreciation of music in adult patients with cochlear implants: A patient questionnaire. *Cochlear Implants International, 4*(2), 85-95.

Ito, J., Takagi, A., Kawano, M., & Honjo, I. (1995). Studies of currently used and experimental cochlear implants. *Acta Otolaryngology, 115*, 493–496.

Huang, N. E., Schen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A, 454*, 903-995.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds revela dichotomies in auditory perception. *Nature, 416*.

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology, 9*, 399-406.