**Theories of audition/the cochlea**

Oxenham et al., 2004:

Recently the debate regarding how pitch is processed in the ear has shifted focus onto temporal explanations. These temporal models pool timing information across auditory nerve fibres without accounting for frequency to place mapping. The goal of the research was to investigate whether temporal representation suffices at accounting for hearing or if tonotopic coding is also necessary in hearing.

Transposed stimuli were used since they can present temporal low-frequency fine structure information to regions of the cochlea tuned to high frequencies.

Results showed that frequency discrimination was worse for transposed tones compared to pure tones. It would seem that the temporal information in the envelope is processed differently from the fine structure. Which is expected as the two represent different aspects of sound perception. However, finding an explanation for this with temporal theories is difficult.

Most sounds in our environment, including speech, music, and animal vocalizations, derive their pitch from the first few low-numbered harmonics (40, 41). These are also the harmonics that tend to be resolved, or processed individually, in the peripheral auditory system (38, 42). Unlike high, unresolved harmonics, which interact in the auditory periphery to produce a complex waveform with a repetition rate corresponding to the reciprocal of the fundamental frequency, the frequencies of the low harmonics must be estimated individually and combined within the auditory system to produce a pitch percept corresponding to the fundamental frequency.

It is suggested that the output of each cochlear filter is expressed as an interval histogram, then its passed through a temporal filter matched to the characteristic frequency of the cochlear filter, producing a spatial representation of the cochlear code.

Erfanian et al., 2016:

The two main theories of pitch perception are place coding and temporal coding. The first revolves around tonotopic representation due to the tonotopic organisation of the cochlea, the other focuses on the timing of action potentials in the auditory nerve where the auditory neurons fire in synchrony (phase locking). The frequency of the missing fundamental entails more than just place-to-frequency mapping.

The normal hearing model performed similarly to when both temporal and place input were present. While place only and temporal only inputs performed differently depending on the types of stimuli. For original vowels, place only inputs performed better then temporal only, suggesting that place cues are more important than temporal ones in making pitch judgments. However, when vowels were high-pass filtered (whatever that means, read methods) temporal cues performed better, implying that cutting the low-frequency regions of the spectrum impaired place cues.

Some long section I didn’t understand because I didn’t read the methods. But it can be concluded that temporal pitch information compensated for the missing spectral information by directing the integrated model towards higher fundamental frequency harmonics.

Zeng, 2002:

Same introduction as before (very well written). Since the two types of pitch coding covary with stimulus frequency, their relative contribution to hearing has been historically debated, still being unresolved.

They are using cochlear implants to differentiate the relative contribution of place and temporal coding.

The results are similar to previous studies on pitch estimation and rate discrimination in electric hearing. The results suggest that the upper boundary for temporal coding is much lower than previously thought (300 Hs instead of 3000Hz). However, this doesn’t mean that cochlear implant users cannot processes higher frequency information, since much higher phase locking has been found in the auditory nerve due to electric stimulation compared to acoustic stimulation. Psychophysical data suggest frequencies as high as 4000Hz can be detected by implant users.

The use of higher than 300 Hz frequency boundary is in line with the capability of phase-locking, however, the high rate stimulation might be due to roughness in timbre rather than highness in pitch.

The most popular autocorrelation theory (e.g Meddis & O’Mard, 1997) combines time and place information. It first computes temporal autocorrelation function in each frequency channel and them sums it across the different channels to create the basis of pitch estimation.

The present data suggest that place coding must be taken into account when modelling complex sound hearing.

Moore, 2008:

When a complex broadband sound is analysed in the cochlea, the result is a series of bandpass-filtered signals, each corresponding to one position in the basilar membrane. This aspect of audition is modelled 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 may play an important role.

The bandpass signal at a specific place on the basilar membrane can be analysed using the Hilbert transform to create the “analytic signal”. Is can be used to decompose the time signal into its envelope (the relatively slow variations in amplitude over time) and temporal fine structure (the rapid oscillations with rate close to the centre of frequency of the band).

Both envelope and fine structure are represented temporally but the latter depends on phase locking to individual cycles of the stimulus waveform. Phase locking goes reliably up to 4-5 kHz with useful information persisting up to 10Hz but the upper limit of phase locking is not known.

Then they go on to explain the role of temporal fine structure in pitch perception. They seem to be proponents of temporal coding.

Cheveign´e, 2004:

Pattern matching models assume that the pitch of the missing fundamental is reconstructed to “fill in” the gaps. There is a close relationship between pattern matching models and spectrum based signal processing.

Temporal and resonance models differ in the time required to made a frequency measurement. For resonance, successive waves cause building up of energy which requires time that varies inversely with frequency resolution. While, temporal models require just enough time to measure the interval between two events.

Autocorrelation and pattern matching are the major models of pitch perception with both extracting the period of the wave but in different ways. Autocorrelation does so directly, while pattern matching does so via a first stage of Fourier transformation. Pattern matching depends on frequency resolution and does not work for unresolved partials. Indeed unresolved partial have weaker pitch, therefore favouring pattern matching. But the pitch does exist even if its weak. Autocorrelation can work for resolved and unresolved stimuli, but there are marked behavioural differences between them suggesting the existence of two mechanisms.

Note: resolved are the harmonics that can be represented somewhere on the basilar membrane (basically they can be explained through place theory) and unresolved are higher-frequency harmonics that are represented together in the auditory filter (basically place theory cant account for them).

Oxenham, 2008:

According to place theory, the pitch of a tone is determined by the auditory nerve fibre which it excites most. Low frequencies cause a pattern of activity along the basilar membrane near the apex and high frequencies near the base of the cochlea. This is called tonotopic representation.

According to temporal theory, the timing of action potential in the auditory nerve determine pitch. Below 4kHz action potentials are more likely to occur at one phase in the cycle of a sinusoid, this is called phase locking and it means the time intervals between action potentials are likely to be multiples of the period of the sinusoid. Information about the frequency of a pure tone is derived by pooling information across multiple auditory nerve fibres (this might be where Fourier transform comes in?).

According to a model using both temporal and place coding theories, the timing information is what is used to derive pitch however it must be presented to the correct place on the cochlea. In practice this would mean that during any point of the hearing process, two points of the basilar membrane will be at different phases in the sinusoidal cycle. Whether two points are in and out of phase with eachother will depend on the frequency of the sound.

The place code theory suggests that pure tone discrimination depends on frequency selectivity, meaning that sharp tuning would result in better discrimination. But pitch discrimination is actually worse at frequencies above 4kHz and pitch selectivity remains constant or even improves as frequencies rise. Regarding temporal coding theories, not much is known about phase locking in humans but research on other mammals would suggest that phase locking only goes up to 4kHz, therefore failing to explain how pitch judgments can occur at even frequencies of 16kHz. One study has suggested that even though phase locking degrades above 4kHz some information can still be extracted in higher frequencies.

Each point along the basilar membrane only responds to specific frequencies and can be represented and a band-pass filter. Therefore, the entirety of the basilar membrane can be represented as overlapping band-pass filters. The relative bandwidths of the filters (i.e the bandwidth as a proportion to the filters centre frequency) decrease with increasing centre frequency of up to 1kHz and then either remain consistent or continue decreasing (depending on how the bandwidths are measured). However, because the spacing of harmonics within a complex tone remains constant on an absolute frequency scale, it is more important to consider how the absolute bandwidth (in Hertz) of the auditory filters varies as a function of centre frequency.

The filter centred around 100Hz has high output when presented with a harmonic complex tone with a fundamental frequency of 100 and partials at integral multiples of that. Conversely, the filter bandwidth centred around 15o has lower output since it doesn’t respond strongly to 100 or 2000 Hz (F0, F1). This type of representation is known as an excitation pattern. Harmonics which produce peaks at filters with centre frequencies corresponding to the lower harmonics are called resolved. At higher frequencies, the bandwidth of the filters exceeds the spacing between adjacent harmonics, meaning that filters cantered between adjacent harmonics respond equally as strongly as filters that are centred around the frequency of the harmonic. At this point, peaks become less distinct and eventually disappear, resulting in unresolved harmonics.

According to models of complex pitch perception based on the excitation pattern concept, just as with place theory, the peaks provide information about which frequencies are present.

According to temporal representation theories, the low-numbered resolved harmonics’ filters centred at one of the harmonic frequencies respond primarily to that single component, producing a response that approximates a pure tone. As the harmonics become more unresolved, they interact within each filter producing a complex waveform with a temporal envelope that repeats at a rate corresponding to the fundamental frequency. The most influential model, based on the autocorrelation function, extracts periodicities from the temporal fine structure of resolved harmonics, and the temporal envelope from high-frequency unresolved harmonics, by pooling temporal information from all channels to derive the most dominant periodicity, which in most cases corresponds to the fundamental frequency.

Then they move on to impaired hearing and cochlear implants. Read maybe.