

Hearing pitch – right place, wrong time?

Chris Plack on competing theories of pitch perception

Pitch perception has been a focus of auditory research for over a hundred years. Yet despite this, we still do not have a clear explanation for how pitch is represented by neurons in the auditory system. Two rival theories have slugged it out over the years: 'place theory', in which pitch is determined by which neurons are active, and 'temporal theory', in which pitch is determined by how the neurons are active, specifically in terms of their temporal firing patterns. Although temporal theory has had the ascendancy for a number of years, recent findings suggest that place theory may be ready for comeback.

Pitch is one of the primary auditory sensations. It is arguably the most important perceptual dimension of music, allowing us to appreciate melody and harmony. In non-tonal languages such as English, pitch variations are part of prosody, and we can stress parts of an utterance by giving particular words a higher pitch. Pitch is even more crucial in tonal languages such as Mandarin Chinese, as it is used to determine word identity. Finally, and perhaps less obviously, pitch is one of the main cues that allow us to separate out sounds that occur together, such as a man and a woman speaking at the same time.

What is pitch?

Sound in general is composed of pressure variations in the air, but sounds that evoke a pitch share a more specific property: they *repeat* over time. Indeed, pitch can be defined as the sensation that is related to the repetition rate of sound waveforms. When the string on a guitar vibrates, it produces a repeating pattern of peaks and dips in pressure. When the string vibrates slowly (e.g. the low string on a guitar) we hear this as a low pitch. When the string vibrates quickly (e.g. the high string on a guitar) we hear this as a high pitch.

The tones made by musical instruments, and by the human vocal apparatus, are complex sound waveforms that are made up of several simpler 'pure tones' of different frequencies. The frequencies of these harmonics are whole number multiples of the overall repetition

rate. For example, if the A string on a guitar is vibrating 110 times a second (110 Hz), harmonics with frequencies of 110, 220, 330, 440 Hz, etc. will be present in the sound wave. The way the magnitudes of the harmonics vary relative to each other determines the *timbre* or quality of the tone. Timbre allows us to distinguish between different instruments (guitar and trumpet) playing the same note, or between different vowel sounds in speech.

Place and time

The human ear has the remarkable ability to separate out the different harmonic components of tones. The spiral cochlea in the inner ear contains a long thin membrane called the *basilar membrane*. Each place on the basilar membrane is tuned to a different frequency, so that when a sound enters the cochlea, the different frequency components cause different places on the basilar membrane to vibrate (with the base of the spiral responding to high frequencies and the apex responding to low frequencies). Hence for a tonal sound, each place on the membrane vibrates at the frequency of the harmonic to which it is tuned.

The sensory hair cells that convert the vibration of the basilar membrane into electrical activity in the auditory nerve are arranged in a single row along the length of the membrane. Since each hair cell is innervated by a separate set of auditory nerve fibres, the frequencies of the harmonics are coded in terms of which nerve fibres are active. This is called 'place theory' as it is the place of activity on the basilar membrane, or the place of activity in the nerve array, that represents the frequencies that are present in the sounds. The 'tonotopic' representation (different frequencies mapped to different places in the auditory system) continues throughout the auditory regions of the brainstem, and is also present in the auditory cortex in the temporal lobe.

However, there is another coding strategy used by the ear. This is based on the *timing* of neural impulses. When the basilar membrane vibrates up and down in response to a sound wave, the tiny 'hairs' or stereocilia on the hair cells sway from side to side. Since the hair cells are only activated when the stereocilia are bent in one direction, the nerve impulses tend to occur at a particular time or phase in each cycle of the sound wave. This process, known as *phase locking*, means that neural firing in the auditory nerve is synchronised to the frequency of vibration of each place on the basilar membrane.

Imagine that the ear is played a tone with a repetition rate of 100 Hz. Let's take

question

How is the pitch of sounds represented by the brain?

resources

www.newmusicbox.org/articles/The-Musical-Ear/
Plack, C.J. (2005). *The sense of hearing*. Hove: Psychology Press.

references

- Attneave, F. & Olson, R.K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, 84, 147–166.
- Gockel, H.E., Carlyon, R.P. & Plack, C.J. (2011). Combination of spectral and binaurally created harmonics in a common central pitch processor. *Journal of the Association for Research in Otolaryngology*, 12, 253–260.
- Meddis, R. & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, 102, 1811–1820.
- Moore, B.C.J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, 610–619.
- Oxenham, A.J., Micheyl, C., Keebler, M.V. et al. (2011). Pitch perception beyond the traditional existence region of pitch. *Proceedings of the National Academy of Sciences*, 108, 7629–7634.

the example of a neuron that is connected to the place on the basilar membrane tuned to a frequency of 100 Hz. This will respond to the *first* harmonic of the tone and will tend to produce nerve impulses that are synchronised to 100 Hz (i.e. the time between impulses will be 10 ms, or integer multiples of this since sometimes neurons do not fire on every cycle). Now let's take the example of a neuron that is connected to the place on the basilar membrane tuned to a frequency of 200 Hz. This will respond to the *second* harmonic of the tone and will tend to produce nerve impulses that are synchronised to 200 Hz (i.e. the time between impulses will be about 5 ms, or integer multiples of this). Hence, the frequencies present in a sound wave are coded, not just by *which* neurons are active, but also by *how* they are active; specifically, by the temporal regularity in the pattern of firing. This is known as 'temporal theory'. In mammals, phase locking works for frequencies below about 5000 Hz (although we don't know for sure

and its perceived pitch, is unaffected by removing the first harmonic, or 'fundamental' component, that has a frequency equal to the repetition rate of the tone. For example, a tone with harmonics of 200, 300, and 400 Hz (the second, third and fourth harmonic of a 100 Hz repetition rate) is heard as having a pitch corresponding to 100 Hz, even though there is no harmonic present at 100 Hz. Similarly, if harmonics of 600, 800, and 1000 Hz (the third, fourth, and fifth harmonic of a 200 Hz repetition rate) are present, a pitch corresponding to 200 Hz is heard.

These observations suggest that pitch is determined by the spacing or patterning of the harmonics rather than by the lowest harmonic frequency present. It follows that the ear must have a mechanism for combining the information from neurons responding to the different harmonics. However, does the ear use information about *which* neurons are active to derive the frequencies of the harmonics, or does

it use the phase-locked temporal firing patterns? In other words, is the correct explanation for pitch perception the place theory or the temporal theory?

Over the last few decades the temporal explanation for pitch has held sway. This view has been based on several converging lines of evidence, three examples of which are provided here.

First, it is argued that our ability to discriminate between two different pitches is much better than would be predicted based on place theory. In optimal conditions, we can tell two tones apart if their repetition

rates differ by just 0.2 per cent (one thirtieth of a semitone) (Moore, 1973). However, the sharpness of tuning of each place on the basilar membrane (the range of frequencies each place responds to) is about 15 per cent of the tuned frequency. Hence, the membrane may not be tuned sharply enough to distinguish frequencies that are so close together. We may need the precise pattern of synchronised firing in the temporal code to provide the sensitivity for these acute discriminations. Above the upper frequency limit of phase locking (thought to be about 5000 Hz), performance is much worse, suggesting that fine pitch discrimination may be reliant on the temporal code.

Second, our ability to perceive musical melodies seems to be severely degraded for

repetition rates above about 5000 Hz (Attneave & Olson, 1971). It is perhaps no coincidence that the highest note on an orchestral instrument (the piccolo) has a frequency of about 4500 Hz. Tunes played with repetition rates above 5000 Hz sound a little strange. You can hear something changing but it doesn't sound particularly musical. So the limit of phase locking seems to coincide with the limit of musical pitch, suggesting that phase locking may be necessary for musical pitch perception.

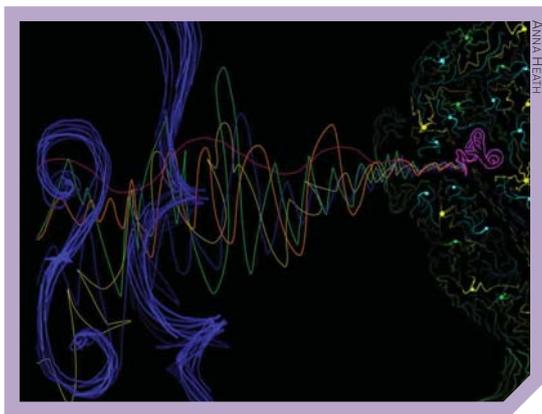
Finally, computer models of pitch processing based on the temporal code have been able to explain a wide range of pitch phenomena (Meddis & O'Mard, 1997), adding credence to the idea that temporal theory may provide a general account of pitch processing.

The story seems quite convincing, and many auditory physiologists and psychologists have turned their attention to where and how the temporal neural code from the different harmonics is converted into a single representation of pitch. However, recent discoveries have cast doubt on the temporal story, and reignited the debate.

Place invaders

Perhaps most dramatic is the recent discovery by Andrew Oxenham and colleagues at the University of Minnesota that it is possible to generate a clear pitch using tones with harmonic frequencies well above the supposed neural limit of phase locking (Oxenham et al., 2011). These authors confirmed that a repetition rate less than 5000 Hz is necessary for a clear musical pitch. However, they also showed that as long as the overall repetition rate is within this range, the harmonic frequency components need not be. For example, they were able to produce a clear pitch corresponding to 1200 Hz with harmonics of 7200 Hz and above (i.e., harmonics of 7200, 8400, 9600, etc. Hz). The pitches produced by these high harmonics supported melody recognition, hence satisfying the criterion for musical pitch. Oxenham and colleagues were careful to control for potential confounds produced by interactions between the harmonics on the basilar membrane.

So why is this bad news for the temporal theory? Well, if the harmonics are above the frequency limit of phase locking, then their frequencies cannot be represented by a temporal code in the auditory nerve. Hence, there is no temporal information related to pitch that could be used by a central pitch extraction process. The results seem to imply that temporal information is not necessary for



The tones made by musical instruments, and by the human vocal apparatus, are complex sound waveforms

that this limit applies to humans). Above this frequency, neurons lose the ability to synchronise their firing to the vibration of the basilar membrane.

To summarise, musical tones are composed of a series of harmonic frequency components and these frequency components are separated out by the basilar membrane in the ear. The frequency of each harmonic is represented in two ways in the auditory nervous system: first, by which neurons are most active and second, by their temporal patterns of firing.

Choosing between the theories

An intriguing aspect of pitch perception is that the overall repetition rate of a tone,

pitch. The alternative is that, for these high frequencies at least, harmonics are represented by the *place* of activation on the basilar membrane and in the neural array.

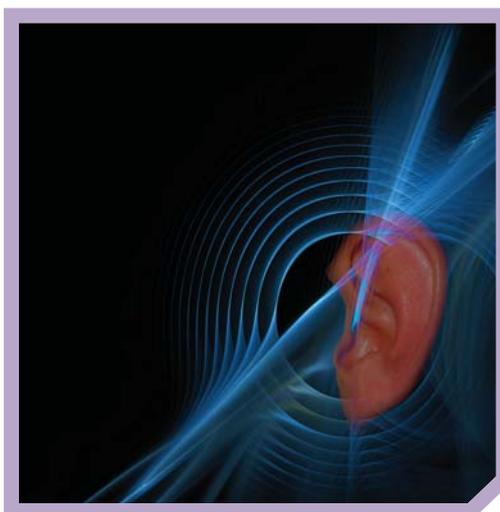
A second recent discovery concerns the location in the auditory pathway at which the harmonics are combined to form a pitch. Unlike the visual system, a great deal of processing occurs in the auditory brainstem before the signal is passed to the auditory cortex in the temporal lobe. From the auditory nerve the signal passes through a number of neural nuclei: the cochlear nucleus, the superior olivary complex, the nuclei of the lateral lemniscus, the inferior colliculus, and finally the medial geniculate body of the thalamus. The precision of the temporal code deteriorates as the signal is passed from neuron to neuron, so that by the level of the inferior colliculus, the maximum frequency of phase locking is several hundred hertz, rather than several thousand hertz.

It has been known for some time that it is possible to hear a pitch produced by just two harmonics that are presented to opposite ears. For example, if a pure tone of 400 Hz is presented to the left ear, and a pure tone of 600 Hz is presented to the right ear, a pitch corresponding to 200 Hz is heard. This implies that harmonics are combined at the level of the superior olivary complex or higher, as this is the earliest stage at which the neural inputs from the two ears are combined.

Hedwig Gockel and colleagues at the MRC Cognition and Brain Sciences Unit in Cambridge investigated this further using an even more esoteric stimulus called 'Huggins pitch' (Gockel et al., 2011). Huggins pitch is produced by presenting the same noise signal (which contains a broad range of frequencies) to both ears except for a narrow frequency band in which the noise is different (decorrelated) between the two ears. A pitch is heard corresponding to the centre frequency of the band. However, the sound in each ear is just a random noise, and it is only when the input from the two ears is combined that a pitch is heard. (I enjoy demonstrating this effect to students using a Huggins pitch melody, as it doesn't seem possible that the tuneless noise heard when the sound is played to each ear separately can be used to produce a recognisable tune when the ears are combined. My aim is to have the first hit single composed entirely of Huggins pitch!)

Again, this combination of inputs from

the two ears probably occurs at the superior olivary complex, but it is thought that the outputs of the neurons that extract the decorrelated frequency band project to the inferior colliculus. Gockel and colleagues showed that not only is it possible to produce a pitch with two Huggins pitch 'harmonics' (for example, decorrelated bands centred on 600 and 800 Hz to produce a pitch corresponding to 200 Hz), but that it is possible to combine a Huggins harmonic and a conventional harmonic to produce a pitch. This suggests that the combination of harmonics in normal pitch perception probably occurs in the inferior colliculus or higher. However, we know that the limit of phase locking is much reduced at this level in the auditory system, so high-frequency harmonics such as 2000, 2400,



We do not know for sure what are the characteristics of phase locking in human listeners

2800 Hz, etc. cannot be represented by a temporal code at this level. A possible implication of the experiment of Gockel and colleagues is that harmonics are encoded using a place mechanism at the stage in auditory processing at which they are combined. This contradicts most temporal models of pitch, although it is important to emphasise that the individual harmonics could be encoded temporally at an earlier stage, and converted to a place code by the level of the inferior colliculus.

A big caveat here is that we do not know for sure what are the characteristics of phase locking in human listeners. For obvious reasons, it is hard to convince an ethics panel that there is scientific justification for opening up the skull and sticking electrodes into the auditory nerve and brainstem of a living human being

(although there are now experiments being planned on auditory nerve measures in patients under surgery for tumours on the auditory nerve). It is at least possible that the frequency limits of phase locking are much higher in humans, both in the auditory nerve and inferior colliculus. If so, then the new results may be accommodated within the temporal account. However, humans would have to be very different from other mammals if this were the case, and it is not clear that phase locking to very high frequencies is even possible physiologically.

Towards a unified theory

The new results cast doubt on the temporal account, and suggest that place theory may well make a dramatic comeback. But will temporal theory now suffer a slow death as we embrace the re-energised place explanation? I think not. There is quite good evidence that temporal information on its own is sufficient for a musical pitch. For example, tones with high-numbered harmonics that are too close together in frequency to be separated by the basilar membrane can still evoke a musical pitch based on the temporal fluctuations in the waveform. More likely is that the auditory brain uses a combination of place and temporal information. After all, there is no reason why the brain should be nice to researchers by using just one type of pitch representation. If both place and temporal information are available, then why not use both?

Determining exactly how (and where) these separate types of information are combined may be the next challenge for researchers.

Vision scientists may laugh at the seeming inability of auditory researchers to provide a definitive physiological explanation of one of the main auditory sensations, despite well over a hundred years of research. However, this is one reason why I find pitch perception such a fascinating field of inquiry. There are still big questions that need to be answered, and this presents opportunities for more groundbreaking experiments in the years to come.



Chris Plack
is Ellis Llwyd Jones Chair in
Audiology at the University
of Manchester
Chris.Plack@manchester.ac
.uk



Essential Supervision Skills

BPS LC Approved Certificate in Clinical Supervision 2 day course

This practical two day course is ideal for those who supervise the clinical and casework of others and or those wishing to train in it.

The course is designed to provide you with an up-to-date theoretical overview of clinical supervision along with its practical application within a range of practice environments. The Certificate draws heavily on psychological theories of therapy, learning and management including the Kolb Learning Cycle and Parallel Process models.

The two days combine an overview of the supervision process with an exploration of the practical problems which arise within it. A particular focus will be on common supervision problems and dilemmas and how they can be successfully addressed.

With lots of opportunities for asking questions, sharing your supervision problems and networking with your colleagues this training is absolutely essential for those already delivering supervision or those wishing to train in it.

Paul Grantham Says:

Clinical supervision is one of those activities that are typically conducted without much training. It is often the case that you're dropped in at the deep end with a supervisee and have to conduct things as best you can. And you make a pretty good job of it! But if you feel you need greater confidence in doing it, or just want a chance to reflect on the latest developments in the field then – this course will be something really special that might interest you.

Why is it special?

Firstly, it is one of the very few short training courses that's received approval from a professional body – The British Psychological Society Learning Centre. That doesn't mean incidentally that it is just for psychologists, but it does mean that we have successfully negotiated and submitted the course for national professional approval from a body that's been in the forefront of supervision developments for the last sixty years.

Secondly...

Read more at www.skillsdevelopment.co.uk

Dates & Venues

LONDON (The British Psychological Society) 5 - 6 FEBRUARY 2013 Fully Booked **FULLY BOOKED**

MANCHESTER (Manchester YHA) 2 - 3 JULY 2013 Fully Booked **FULLY BOOKED**

LONDON (The British Psychological Society) 9 - 10 JULY 2013 **Places available**

BIRMINGHAM (The Ibis Hotel) 19 - 20 SEPTEMBER 2013 **Places available**

Book before 1 January 2013 and save over £70 off the full course price!

www.skillsdevelopment.co.uk