An investigation on similarities between pitch perception and localization of harmonic complex tones

Jon Allan
Abstract

This paper investigates two types of human auditory perceptions, localization and pitch perception, to see if there could be any common mechanism underlying both abilities. A brief description of the theories will be listed. From those theories it will be shown that, for lower frequencies, both perceptions seem to be based on timing information from the hair cells in the cochlea. One experiment will also be outlined and performed to investigate our ability to localize harmonic complex tones, depending on how many partials the tones are built of. The results from the experiment are compared with earlier reported experiments on pitch perception and some interesting similarities are enlightened that implies there could be a common mechanism underlying both perceptions. However, more subjects are needed to be able to draw reliable conclusions.
# Table of contents

**INTRODUCTION** .......................................................................................................................... 4

**PITCH PERCEPTION** ..................................................................................................................... 5

  - Place Theory .......................................................................................................................... 5
  - Frequency Theory .................................................................................................................. 5
  - Pitch Perception .................................................................................................................... 5

**LOCALIZATION** ........................................................................................................................ 8

**SIMILARITIES BETWEEN PITCH PERCEPTION AND SOUND LOCALIZATION** ............. 11

**EXPERIMENT** ............................................................................................................................ 14

  - The Hypothesis .................................................................................................................... 14
  - The Idea ................................................................................................................................ 14
  - Outline .................................................................................................................................... 14
  - The Subjects .......................................................................................................................... 15
  - The Test Tones ....................................................................................................................... 15
  - The Procedure of Replaying .............................................................................................. 16
  - Preliminary Study .................................................................................................................. 17
  - Equipment Used .................................................................................................................... 17

**DATA ACQUIRED FROM THE LISTENING TEST** ................................................................. 18

**ANALYSIS** .................................................................................................................................. 20

  - Overview ................................................................................................................................ 20
  - Outliers .................................................................................................................................... 20
  - Calculation and Comparison of Standard Deviations .......................................................... 21
  - The Results from a Pitch Identification Experiment ......................................................... 22
  - Comparison of Results between Two Experiments ............................................................ 22
  - Test of Significance ............................................................................................................... 24

**RESULTS** .................................................................................................................................... 24

**DISCUSSION** ............................................................................................................................... 25

  - Reflection on Method ........................................................................................................... 26
  - Suggestions for Further Work .............................................................................................. 26
  - Further Reading ..................................................................................................................... 27

**ACKNOWLEDGEMENTS** ......................................................................................................... 27

**REFERENCES** .............................................................................................................................. 28
Introduction

According to Handel, perception is how the individual interprets all information from the senses into an understandable world of object and events [1]. Perception is something that is essential for the human to understand what she hears or sees etc. to be able to make the right decisions on how to act. As concerns auditory perception, it supplies information about the positions, states and properties of sounding objects around us through perceptual qualities such as pitch and timbre. However, Neither pitch nor apparent position can be directly related to specific physical properties of a sound wave or wave front. The way we perceive these qualities cannot be accurately modelled and calculated by a machine. Based on our knowledge so far, simulations in computers can do calculations that resembles our perception, but there is still many situations where the computers go wrong.

Research on perceptions is difficult. Percepts are often ambiguous, they vary from person to person, and even for the same person they may vary from time to time depending on different circumstances. One problem with research on human perception is that for methodological reasons only a few factors can be investigated at a time, if one should be able to draw statistical conclusions. But perception seems to be based on complex mechanisms and is likely to depend on several factors.

In most research papers in the psychoacoustical field of research, the perception of pitch and localization have each been investigated in isolation, and different theories could be found in the literature for each of these areas. But are these two percepts really totally isolated phenomenon? In Handel [1] states: “Listening is an active process. Auditory information is context dependent; the significance of any part of the signal varies as a function of the other parts of the signal. This makes it unlikely that perception is based on the value of any single attribute”. In this essay, I’m trying to find out what similarities there could be between the mechanisms behind our ability to detect pitch and our ability to locate sound sources in space. Are there experiments showing that pitch could affect the perception of localization and vice versa? Could the two mechanisms be connected or have some common basis in our brain? Could pitch perception be an important part of our ability to locate sounds? My belief is that the two percepts have a relationship that is of interest to investigate. If it could be shown that perception of pitch and of location have some essential traits in common, this information would be useful for the physiologist in studying and understanding those neurological structures in the brain which belong to the sense of hearing.

Some clues to support the issue are worth mentioning. A sound engineer has often encountered the following phenomena: A pure sine wave is hard to localize. An interesting point is that a pure sine also gives a quite diffuse sensation of pitch. Among other aspects it has been shown to be dependent on the sound level. Thus the pitch for a sine above 2000Hz tends to rise when the sound level is increased and a frequency under 1000Hz tends to give a lower pitch as the level is increased [11]. As said above, our auditory perception is essentially based on defining objects and correlating sound sensations to these objects [1]. A sine is not a very defined object. The non-technician perhaps remembers those old cell-phones with the simplest ring-sounds, which were hard to locate. Everyone tended to look in different directions when a telephone rang in a public place. Today there is no such problem as the signals are more complex. Some people have been reported to describe the sound from birds
as continuously changing position, even if the bird sat still. Could this be coupled to the different pitches in the bird's song? An old issue was a finding that a higher pitch tends to be perceived as coming from a higher vertical place than a lower pitch. Could this be confirmed? To be able to draw some conclusions about the issue here raised, we need first of all to look a bit closer into the theories of pitch perception and localization.

**Pitch perception**

Pitch could be defined as a perceptual characteristic of a sound, which is described as “high” or “low” and among other things determines its position on a musical scale. One important information, extracted early in the perceptual process, is the frequency content of the sound wave. You could say that the ear performs a kind of Fourier analysis. Two main theories of how the inner ear may extract frequency information are still applicable today. Each theory has its strengths and weaknesses, and the common opinion is that both theories are to some extent true and back up each other. In other words, the human brain has different approaches to solve the same problem, and the method that is most relevant in the particular case is the method used. Here follows a recap of the two theories:

**Place theory**

When a sound hits the ear drum a pressure wave is set up in the cochlea and makes the basilar membrane vibrate. Due to the different characteristic of the basilar membrane at different places one could expect that there exists different resonant frequencies along it. A particular frequency should have its particular place, where the membrane has its largest amplitude variation. This has been shown to be the case to a great extent for high frequencies but for low frequencies the whole membrane is moving, especially at high intensity. Hence our high resolution in detecting different frequencies in the lower regions cannot be satisfactory explained by the place theory.

**Frequency theory**

The concept is that the frequency in which the membrane moves up and down is directly coded as timing intervals between the nerve pulses. For example each time the membrane reaches its top the nerve fiber in that position fires. This theory could give a precise frequency information, but gives rise to problems when many frequencies occurs at the same time. However, it has been empirically proved that this kind of timing information from the nerve cells does exist. This information will further on be referred to as *phase information*.

More on the construction of the ear and the two theories can be found at [1, p.461].

**Pitch perception**

When frequencies in the received sound have been detected, the question arises, how several frequencies together produce the perception of pitch. One explanation is that harmonic frequencies produce a repeatable pattern. The periodicity of the combined waveform is the frequency to which all the combined frequencies are multiples. This
frequency or fundamental is often matching the perceived pitch. “Most musical instruments have a clear pitch that is associated with the periodicity of the sound they produce” it is stated in [4]. The periodicity theory claims that that the hearing sense is able to recognize the repeatable pattern and derive the frequency of it’s repetition rate. Exactly how this would be done is still unclear, but one idea is that the hearing sense recognizes the amplitude pattern of the waveform to find out when it repeats itself. Anyway, some sort of time analysis mechanism that uses the phase information from the cochlea would be needed to backup the periodicity theory.

An important phenomenon is that a pitch corresponding to that fundamental is perceived even when the fundamental itself isn’t represented in the spectrum. This conforms well with the perception, because we can still hear the pitch even if the fundamental is physically absent! This perceptual phenomenon is often referred to as “pitch of the missing fundamental”, “virtual pitch” or “musical pitch” [8][10]. It conforms well with the fact that the periodicity can still be seen even when the fundamental is missing (see Fig 1a). Figure 1b shows which partials the above waveform is built of, and the dashed line represents the fundamental frequency. One example when this phenomenon can be noticed in everyday life is when one listens to music through band limited channels, like a small transistor radio or a telephone. One can still follow melodies of low-pitched passages without ambiguity, despite the absence of acoustical energy at the fundamental frequencies of the notes. More on the interesting phenomenon may be read in [8] where different explanations and models are reported.

![Figure 1a](image)

Here, five partials are blended, 450, 600, 750, 900 and 1050 Hz. They are all multiples of the same frequency, 150 Hz, which clearly is seen as a periodicity in the waveform.
There are however exceptions when the periodicity theory will not work. The clarinet for example generates only odd harmonics (at least they are very dominating). Here our perception won’t match the repeatable pattern. In experiments with only odd harmonics, in many cases we hear a pitch that is one octave higher than the frequency of the repeatable pattern [4]. Here the pitch perception rather seems to be based on the distance between the harmonics instead (see Fig. 1c).

In 1924 H. Fletcher showed that at least three successive partials are needed to create a sensation of pitch [4]. Terhardt defined this perceived pitch as *virtual pitch* as opposed to when one hears a single frequency which he called *spectral pitch*. In [4], it is further stated that the sensation of pitch gets stronger as the number of partials are increased. The phenomenon of virtual pitch is present for lower frequencies up to 1000Hz. Above that it’s not possible to hear the pitch without the fundamental.

Houtsma and Smurzinski (1990) explored this further and showed that pitch sensation depends on waveform or periodicity at lower frequencies and some other mechanism at higher frequencies. This other mechanism conforms well with the physiological studies and the place theory. The nerve cells can’t fire at this high speed intervals, but instead the basilar membrane is more efficiently tuned to different frequencies at different places.
Even in the case when the frequency spectrum of the sound does not show the regular pattern of multiplicities of a fundamental some kind of pitch is often perceived. For instance when the harmonics are randomly chosen. Another case is so called repetition pitch, which is heard when one or more delayed repetitions of a sound combine with the original sound. This pitch corresponds to the inverse of the time delay and is audible for time delays ranging from 1 to 10 ms [10].

Localization

By localization in this context, is meant the ability to make, in the mind, a spatial representation of sounding objects around us in the real world. This ability is quite amazing, because the only information available for doing this is a neural representation of the movement of the two eardrums in our ears. Vision could be of help, but localization will work well without vision too. The task will be especially intricate when several sound sources from different directions blend together. The mind has to understand which components of the sound wave belongs to which object, and assign that object a position. Even if we move our head, the representation of the world outside seems stable.

Even if it is possible to localize a sound to some degree with only one functional ear this ability is vastly enhanced by the use of two ears. Given the neural information from the two cochleas it should be possible to calculate time differences and level differences for different frequencies. These interaural time differences occur because one ear is closer to the sound source than the other. The largest time difference that could occur is 0.65ms (interpreted from a diagram in [1]). This corresponds to a sound coming straight from the left or the right side of the head. Our hearing seem to have a good delay function to be able to detect when signals from the two ears match, and from this delay calculate the incoming angle [3]. In addition to this, higher frequencies (>1500 Hz) will be filtered by the head and arrive at the farther ear at a lower intensity. The shape of the head, the outer ears, neck and shoulders will also affect the levels of different frequencies depending on the direction of arrival of the sound. These forms will reflect and diffract different frequencies of the sound waves in unique ways depending on the direction of the arrival of sound. It is this property that makes it possible to distinguish a sound coming from the front from a sound coming from the back and also to register from which height a sound is coming from. When doing computer models the functions that transforms the three-dimensional direction of a sound to a two-channel simulation of the movement of the eardrums are often referred to as HRTF:s, Head Related Transfer Functions. By using a Stereophonic Head for recording, the three dimensional space is also captured to two channels in a similar manner to our own heads. These recordings may then be played back in headphones to get a feel of a true three-dimensional environment.

Reflections from walls, floor and ceiling could be of additional help for localizing a sound, but could also make the task more difficult. But we will leave this aspect as we soon will narrow down to the parts of the localization problem that will be of interest for this essay. The distance to a sound source is another aspect of localization that will fall out of the scope of this essay.
Something more should however be said about interaural time differences. The onsets and offsets of sounds are especially efficient cues for localization. However, even for continuous sounds there exists important timing information. There is strong evidence that regular amplitude oscillations, such as those created by beats, are also used for localization [1]. The definition of beats is as follows: In the simplest form, when two pure frequencies (sine waves) are superimposed, there will be an amplitude oscillation with a frequency that is the difference between the two mixed frequencies (see Fig. 2). Every maximum in this oscillation corresponds to a beat. If you mix several frequencies that all are multiples of a common frequency you will see a regular amplitude pattern that will repeat itself with the common frequency. In this case you also have amplitude beats (see Fig. 3). Depending on the phase relationship between the inherent frequencies, the beats are not always as clearly seen in the combined waveform as in this illustrative example. But the phase relationships does not seem to affect our ability to detect the regular amplitude pattern. It works equally well irrespective of the phase relationships. Then the timing differences of the corresponding beats between the ears will contribute to the information used for localizing continuous sounds.

It is this last aspect of localization that will be of interest for this essay. That we are able to detect a regular amplitude pattern and compare the patterns from the two ears to find the time difference. To be able to investigate this aspect alone, we will create a test environment for the experiment (described later on) where all other influential conditions of importance for localization will be eliminated.

By using headphones a lot of factors that are hard to control will be removed. It will remove any reflections from walls, roof and ceiling or other objects. The use of headphones will also remove the effect of head filtering, leaving a total control over time and level differences to the test material played to the subject. On the other hand, in using headphones the perception of localization will be a bit artificial. At least when not using HRTF:s or reverberation. Instead of hearing the different sounds coming from directions around you, you will hear the sound like they were inside your head. The dimensions of distance, height, forwards or backwards are no longer there. The only dimension that is left is the left to right dimension. It is still however easy to localize different sounds along this one-dimensional scale, and sound engineers in particular are used to listen to this artificial space. The term localize usually implies a three-dimensional percept of a sound source, but from now on, for this essay, the term will be used in a narrower sense and only refer to the left-right dimension of localization. The term location will be used in a similar manner for a perceived position of a sound source along the left-right axis.
Figure 2

Sine width a frequency of 200 Hz

Sine with a frequency of 230 Hz

Combination of Sine with 200 Hz and 230 Hz. The frequency of the beats is 30 Hz.
The difference between the two combined frequencies.

Figure 3

The sum of the frequencies 180, 210, 240, 270 and 300 Hz. They are all multiples
of the frequency 30 Hz, which is in this picture also seen as a beating pattern.
Similarities between pitch perception and sound localization

From the above descriptions of the two perceptions we can derive the following similarities:

• Both pitch and localization are involved in our perception of the environment. This may seem unnecessary to state, but considering the quotation at the beginning of the introduction, it means that they already have a lot in common:

  - Both percepts are bound to objects and events [1]. The whole idea of a perception is to give us an understandable “map of the world”. Neither location nor pitch would be ecologically useful unless coupled to an object. (Of course they are musically useful, even if not related to material objects, but that is another story, maybe.)
  - Both elements have gone through some neural processing. Perception is the result of weighing several information paths together to form a usable percept. This means that both pitch perceptions are probably formed at a high neurological level in our brain (see Fig. 4). At least at a higher level than the original information from the cochlea. Maybe both percepts are formed in the auditory cortex, the highest level of the auditory sense.
  - They can both be ambiguous or vague and differ from time to time or from person to person. At least in experimental situations were the limits are pushed, often by reducing the amount of information that the subject can use to form a percept.
  - It’s hard to make a correct computer model that will make the same judgments as our hearing sense does in every situation.

• Both abilities have strong correlation to time and patterns.

  By comparing Picture 1 from the description of pitch with Picture 3 from the description of localization you can see the resemblance. In both cases the ear seems to be able to extract the amplitude pattern. In the first case to be able to find out when the wave form repeats itself, and in the latter to grasp the time difference between the ears. The concept of “periodicity” from the chapter on pitch perception and the concept of “amplitude beats” from the chapter on localization are both explained as a lower frequency combination of higher frequencies. Exactly how this analysis is done is not known for certain, but the interesting part is that it could be the case that the same information is used for both purposes. It would feel like a waste of neurons to have two different approaches to extract the same necessary information used for two different purposes. On the other hand, if two different approaches are needed and do exist, these two approaches would be helpful for both pitch detection and localization. In either case we should have interesting coupling.
The black arrows represent some of the neural pathways belonging to the auditory sense. The circles represent “relay stations” where different kinds of nerve information and sometimes information from the two different hemispheres interact. The idea of this picture is to show that the expression “high neurological level” corresponds to places where nerve signals have passed several relay stations since the origin of the information from the cochlea. Percepts like pitch and localization probable occur in the higher regions, for example in the Auditory cortex. To get a more thorough understanding of the physiology behind the auditory sense, one can read the chapter “Physiology of Listening” in “Listening” by Handel [1]. Here the author summarizes and describes the research that have been made in the area in a comprehensive way. This picture is made with a similar picture from the book as a starting point.

To this we can add the following consideration:

- Physiological properties.
  
  One approach other than to examine humans subjective descriptions is to look at the physiological parts to see if one can find similarities. The first obvious similarity is that pitch as well as localization is extracted from the same waves described by the motion of the oval window. In the next step different nerves will fire as the basilar membrane moves and both perceptions will have access to the same neurological information from the cochlea to begin with. For
frequencies up to 4000 (or maybe 5000 Hz at most) there has been shown that nerve cells fire in sync with different sine-components of a wave (although with different limitations). As mentioned earlier you could say that there exists a strong phase information. This information is a very important factor for locating continuous sounds. Jan O. Nordmark, for instance, concludes that “Localization studies with low-frequency tones definitely show that the temporal pattern is preserved” [7]. Handel writes “There is strong evidence that regular amplitude oscillations, such as those created by beats, are used for localization” [1]. It is this property that makes it possible for the brain to calculate a timing difference between the signals from the two ears, and with this timing difference relate to an angle. At least for sounds with low frequencies where this phase information exists -- the phase information however is also essential for pitch detection. Without this phase information it would not be possible to extract the lower frequencies which are needed. So the phase information from the cochlea seem to be used by both perceptions.

• Other reported experiments

In [7], several experiments are described which all point in the direction that there probably is a common underlying mechanism to both pitch perception and localization. Nordmark describes this as a time-interval measurement mechanism that will use phase information as its input. He further states that this mechanism can explain the identification of a pure tone, as well as the detection of periodicity in a waveform and moreover account for pitch sensations in non-periodic sounds.

In one experiment [7], a delay was induced between two identical pulse trains. The accuracy in determining the pitch of the pulse trains when led to the same ear was compared with the accuracy in localizing the pulse trains when fed to different ears. Different degrees of randomization was added to the timing of the pulses to see how the least discriminable pitch and the least discriminable location was affected. The result was a strong match in behavior of the two tasks and N. states: “Nevertheless, the similarity between the mechanisms underlying pitch and lateralization seems to be established beyond doubt”. The “nevertheless” was alluding to the fact that the experiment had very few subjects, and more data was needed to accept the relation.

For higher frequencies pitch detection solves the time-resolution problem with the help of the place theory. Luckily our localization ability has a solution in that case too. It’s only at the higher frequencies that the head and outer ears will affect the intensity of frequencies depending on the direction of the sound. So for higher frequencies probable localization and pitch use different approaches.
Experiment

The hypothesis

The hypothesis I want to put to the test in this study can be stated as follows: There is a close resemblance between the functional mechanisms behind human perception of pitch and localization, indicating that partly the same physiological system may serve both kinds of perception.

The idea

The method has been to compare the results of different listening tasks to see if there is a pattern that is alike between pitch perception and localization.

As mentioned above, in the section on pitch, three consecutive partials are needed in the composition of a tone for a virtual pitch to emerge. The present experiment poses the question: Is this also a criterion for localizing a tone. Of course all sound is localizable in one way or another, but here the expression will have the following restraints:

1) That all subjects locate the tone to the same location (with a reasonable amount of spread).
2) That the location could be related to either the time difference between the ears or to the phase shift that belongs to the perceived fundamental.

It is known that the sensation of pitch gets stronger – or more well-defined, less vague -- when the number of partials are increased. Will it also be easier to locate a sound with many partials, than to locate a sound with only a few partials?

Accordingly, the idea of the experiment here reported has been to find out how accurately the subjects are able to locate different tones depending of how many partials the tones are built of. If the results show that more partials give a more defined location of the sound, then this result resembles the results from experiments concerning pitch perception.

Outline

A number of tones consisting of harmonic partials without fundamental were played to subjects in headphones. Their task was to judge the location of each tone. Headphones was used to eliminate factors that influence the perception of location other than the phase information. One reason for building the tones without fundamental is that the results will be compared with the results from experiments on pitch perception with “missing fundamental”, so we need to have as similar test conditions as possible.

The subjects judged nine different tones. The order of the tones was randomized for every subject. After the series the subject will judge the nine tones once more, but this time with a new randomized order.
The subjects

All subjects were chosen from the Sound Engineering Program, at the Institute of Music and Media at the Luleå University of Technology. When the students applied to the school they went through a hearing test that concluded they had no hearing loss. They also did a musical test, that insured they are able to determine the pitch of a tone as well as detect which tone is the lower one of two tones. A total of 16 subjects has made the test. Four women and twelve men. To get an equal number of women and men was strived for, but would unfortunately have lessened the number of subjects available considerably.

The test tones

All tones had a total length of three seconds. They had a fade-in and a fade-out of one second. This is to reduce the onset and offset cues. The fades were applied after the time delay described below. The fundamental frequency for each tone was 150 Hz. The frequency was chosen not to be too high, since the partials will fast build up to higher frequencies where time information is poor. It will not be too low either, in which case it would be hard for a virtual pitch to emerge [4].

Three sets of tones were used. One set with tones created with 2 partials, a set with 3 partials and a set with 5 partials. All tones had the third partial as their lowest frequency and then build up with consecutive partials upwards. The fundamental and its octave was thus removed in all tones. In each set there were three versions of each tone. All tones were fed to both channels to the headphones but one of the channels (at random) was delayed using one of three fixed times, 0.1 ms; 0.3 ms or 0.65 ms. The three sets each with three different time delays make up to the nine different tones used. Table 1a shows an overview of the test tones used for the experiment and Table 1b shows the frequency content in the test tones depending on how many partials the tones are created of.

As stated above, which channel is delayed is randomized, but once randomized it will not change between the repetitions and will be the same for all groups. All tones will be normalized to get the same peak level. A listening level of 58-60 dB dBA was used throughout all tests. The listening level was established during preliminary tests. The measurement technique of this level is a bit uncertain though. To get a correct reading of the level a dummy head with calibrated measurement microphones would have to be used, and this equipment was not at hands. This is also the reason for the interval of the level readings. The measurement was made with a Monacor SM-2 decibel meter, with fast reading setting.

The level of the different frequencies building up the complex test tones were established as follows. The level of the lowest order partial, number three, was set to –9dB (peak value) below the digital zero. For each higher order partial the level was decreased with 1 dB. Thus the level of partial four was –10db, the level of partial five was –11dB etc. The reason for having this slope in level is that we seem to integrate the partials easier, to hear them as a whole, instead of the higher partials “breaking out” of the complex and be perceived as single frequencies [1]. The reason for this is probably that the character of a timbre is resembles more those timbres that exists
Table 1a. Overview of timbres used in the experiment

<table>
<thead>
<tr>
<th></th>
<th>2 partials</th>
<th>3 partials</th>
<th>5 partials</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 ms</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>0.3 ms</td>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>0.5 ms</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

Delay time between left and right channel

Table 1b. Frequency content in the test tones

<table>
<thead>
<tr>
<th>No. partials</th>
<th>Frequency content in Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>450, 600</td>
</tr>
<tr>
<td>3</td>
<td>450, 600, 750</td>
</tr>
<tr>
<td>5</td>
<td>450, 600, 750, 900, 1050</td>
</tr>
</tbody>
</table>

Table 1a shows an overview of the test tones used in the experiment. The tones are labeled one thru nine, and by looking at the labels for each row and column, one can see how many partials the tone is created of, and which time delay between left and right channel is used.

Table 1b shows the frequency content in a tone depending on how many partials it is created of.

In nature and in harmonic acoustical instruments. Pink noise is an example were the energy drops 3 dB per octave band, and this energy distribution seem to be more natural to our ears than white noise where all bands has the same energy level. The amount of decrement per partial was tested in the preliminary tests, and this method of decrement gave complex tones that sounded more collected than if the partials had been mixed with equal level. The difference in level between partial three and six is 3 dB, and the interval between the partials is one octave, so this kind of slope resembles that of pink noise. After all complexes had been created, they were all normalized so there respective peak levels would become –10 dB. This would ensure that all tones would play with the same level (SPL) in the headphones and thus minimizing the risk that the level of the tones would affect the test results.

The procedure of replaying

At any time during the tasks below the subject can run the tone as many times he/she wants to check his/her setting. When the subject is satisfied with the setting, a new tone will be presented and the procedure will be repeated. After half the tones are gone through there will be a five minutes break. The total test was in most cases between 35 and 50 minutes.

For each tone, the subjects were asked to find the location of that tone. This was done by letting the subject control a sounding pointer. The pointer can be turned on and off as many times the subject wants, but the subject is asked never to have the pointer on simultaneously with the test tone. The pointer is made in a Pro Tools system. The same constructed waveform of pink noise is applied to both ears (channels). Both channels were delayed 100 samples (in 44.1 kHz). The delay in the right channel will then be controlled by the subject with an external knob (In this case one of the knobs in Pro Tools Control24. The Option key must be held down to get high resolution in the knob). This will cause a shift in the position to the left or to the right with an
amount that the subject will control. This “time pan” is thus created exactly the same way as the pan of the tones. Noise as a source is chosen because it will not be confused or interfere with the tones that shall be evaluated. The resolution of the panning will be the amount of samples shifted.

When the subject is pleased, he/she is asked to write down a number, seen on a computer screen. The number corresponds to the position of the pointer. By the way, the visual indicator on Control24 is obscured so as not to affect the subject. The subject is also asked not to look at the computer screen until finished with the evaluation of the tone.

Preliminary study
Preliminary experiments were performed before the real study. In the test experiments pure tones (sine-waves) were tested to look for the upper limit where phase information could be used for localization. The same frequency was played to both ears and a time delay was induced to one ear. My conclusion was that above 1200 Hz it was very hard to perceive any shift in position and above 1500 Hz it was impossible. (different delays were tested). This seem to fit well with information read in [1].

Therefore I concluded that I wanted to use only frequencies below 1200 Hz. The use of lower frequencies only will also lessen the risk of giving other location cues related to HRTF:s. Even if the use of headphones probable will eliminate the impact of the head, it could still be possible that the hearing sense reacts to this higher frequencies in a way that I have no control over.

I wanted to remove the second partial as well as the fundamental in the timbres. Many experiments remove the second partial because it’s easy to confuse the octave with the fundamental, and it’s harder to know if it’s really the fundamental that the subject is perceiving. So, starting from the 3:rd partial and building upwards, the last partial in the 5-partial tone will be the 7:th partial. The fundamental frequency could therefore not be to high or the highest fundamental will exceed 1200 Hz.

I tried the full experiment myself with both 150Hz and 400 Hz as the fundamental. The 150Hz-tones gave much more interesting results. I discovered a greater difference between the test tones in my ability of localizing them with this fundamental frequency than for the higher one. The 400 Hz:s highest partial will be 2800 Hz, and I believe that this is the reason why it doesn’t work very well.

I established a listing level that was comfortable. When the listening level got higher, the combination tone of the two-partial tones was heard clearly. I wanted to reduce the listening level, so this combination tone was close to not perceivable.

Equipment used
The tones were played from a Pro Tools system with Mbox. The signal was transmitted digitally to another Pro Tools system where the pointer was created and where the signals where mixed. The session was made in 24 bit with a sampling
frequency of 44.1 kHz. A Control24 was used as the subjects control surface where one of the knobs was controlling a Timeadjuster plugin in the latter Pro Tools system. All sound were outputted through a Pro Tools HD 192 Interface. Electrostatic headphones were used with the brand Stax SRM-Xh.

Data acquired from the listening test

The expression “time group” will hereafter refer to the group of test tones that share a common time delay. The time groups will be labeled 0.1, 0.3 and 0.65 depending on the time delays used when constructing the tones, and each group will consist of three test tones (with three different sets of partials).

The data acquired from the listening tests is shown in Figure 5 as normal probability distribution plots to get an easy and understandable overview. The plots were made in Excel by using the Normdist-function. This function uses the mean and the standard deviation from a set of data to calculate the normal probability mass function. One property of the resulting curves is that the area beneath each curve is always 1. This means that a curve with a narrower spread also will reach a higher maximum. Therefore it’s easy to compare the curves by their respective height to see which curve has the more narrow spread.

This representation of the data demands that the data really are normally distributed. To confirm this, the distribution of the data for each test tone was checked for normality in the Statgraphics program with the function “Tests for normality”. It was confirmed that the data were normally distributed for all tones, or put more accurately: It could not be proven for any of the tones that they would not be normally distributed. The program was set to perform a Shapiro-Wilks test with 5% as the significance level. In the Users manual for the program one can read that this function puts the hypothesis (in this case the null-hypothesis) that the data comes from a normal distribution into trial. If the test results in a p-value below 0.05, this will mean that one can reject the null-hypothesis and that this assumption will have a 5% probability of being wrong. For the tests that were made, all p-values exceeded 0.05 which resulted in the null-hypothesis not being rejected. It is therefore likely that the distributions are normally distributed with a probability of 95% for each tested set of data.

The fact that the data are normally distributed is also a criterion for performing the F-tests in the statistical analysis below. The normal distribution plots as well as the above described tests for normality were made after “outliers” have been removed (see below). Three different plots each covers the representation of one time group. This will facilitate the comparison of the difference in spreads between the tones according to how many partials they are constructed of.
Figure 5

Normal distributions, test results for 0.1 ms

Normal distributions, test results for 0.3 ms.
The diagrams shows the normal probability mass function for the test results for the different tones. The area beneath each curve is always 1, which makes it easy to compare the curves to see the differences in their respective spread. It is directly seen that the red ones corresponding to 5 partials, are the most well-defined – or as a matter of fact in the 0.3 group actually has somewhat broader spread.

**Analysis**

**Overview**

The idea of the analysis is to look at the spreads of the subjects judgment of location of the different tones. A large spread will imply that it is harder to locate that tone or that the perceived location of the tone is more ambiguous. The spreads of all the subjects judged locations for each tone was calculated with the standard deviation formula. The ordering of the spreads within each time group was investigated to look for similar patterns between the groups. The differences between the spreads within each time group was also compared with the F-test to check whether they are statistically significant.

**Outliers**

Some of the results were far out from the distribution of the rest of the results for a tone. These outliers had to be dealt with because there were not many subjects in the study, and the large deviation of the outliers would make a great impact on the statistical analysis. In [9] an accepted method for doing this is explained. All results that differ more than two standard deviations from a groups mean may be excluded. However, when this has been done the process must be repeated iteratively until all values are within the plus minus two standard deviation bound. It is important that
this procedure will be done to all values in all groups that will be analyzed. When this method was tried however, a lot of data was removed. Even results that in the beginning seemed to be meaningful and part of a normal distribution. Therefore the method was tried again on the original data, but this time with a harder criterion on the outliers. Only outliers that differed three standard deviations from the groups mean were removed. The process was repeated iteratively and the same process was used on all data. More data was in this way saved for the analysis and the removal of the extreme outliers would give a better foundation for the statistical analysis. A total of 14 data points were removed from a total of 288. The reason that these outliers occurred, could be the following: When the subject reached the extreme settings with the pointer, a time-delay of plus or minus 3 ms was created. This is beyond the limit of the time-difference that could occur “in real life”, with a real sound source and our hearing sense could not possible interpret this large delay as any meaningful information for location. Therefore the pointer does not accurately represent a location based on the time difference and instead uncontrollable effects could occur. Perhaps a limitation of the extremes in the positioning of the pointer would have solved the described problem, but it would be difficult to implement such a limitation to the experimental gear that was at hand.

Calculation and comparison of standard deviations

The results of the calculations of standard deviations for the different tones, after the outliers were removed, were as follows:

Table 2. Standard deviations

<table>
<thead>
<tr>
<th></th>
<th>0.1</th>
<th>0.3</th>
<th>0.65</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.92</td>
<td>5.13</td>
<td>11.70</td>
</tr>
<tr>
<td>3</td>
<td>3.60</td>
<td>3.90</td>
<td>9.55</td>
</tr>
<tr>
<td>5</td>
<td>3.02</td>
<td>4.35</td>
<td>6.93</td>
</tr>
</tbody>
</table>

no. partials

If one arranges the test tones within each time group according to its standard deviation, one gets the results as shown in the table below. The expression “Smallest spread” conforms to the tone with the smallest standard deviation. “Largest spread” conforms to the largest standard deviation.

Table 3. The order of the spreads

<table>
<thead>
<tr>
<th></th>
<th>0.1</th>
<th>0.3</th>
<th>0.65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smallest spread</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Largest spread</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

First of all it’s easy too interpret from Table 2, that it is harder to localize the tones to the sides (time delay = 0.65 ms) than around the median (time delay = 0.1 ms). This was not of interest for the study however, but the results fit well with other studies that have established that we have a poorer resolution in localizing a sound to the sides than we have to localize it around the median [3]. The interesting part here is to
see if the tones with more partials are easier to localize than those with only a few. By looking at the standard deviations in Table 2 and the ordering of the spreads in Table 3 it seems likely that this could be the case. The only exception to the ordering where the 5-partial tones have the smallest spread and the 2-partial tones have the largest spread is the order of the values in the area that is shaded in colour. Considering the consistency of the rest of the results, it seems unlikely that this exception should be a real effect. It is probably the result of experimental errors and that if more subjects were to do the test, giving more data to the study, even the order of the 0.3 ms. delayed test tones would be the same as for the other time groups.

For the following figure (Fig. 6a), we define the accuracy in localizing a tone as being the inverse of the spread of the test results for that tone. The accuracy is calculated as \(1/(\text{standard deviation})\) for each tone. The different time groups are interpreted separately, as shown with the help of different colours. For each of the time groups, there could be seen a rising trend as the number of partials increases from two to five.

The results from a pitch identification experiment

In Fig. 6b the results from a pitch identification experiment by Houtsma and Smurzynski are shown [8]. The experiment tested the ability to identify a correct pitch in a tone depending on the number of partials (in the figure labeled components) the tone was built of. The graph shows that the ability to identify the correct pitch increases as the number of partials go up. Especially in the interval from two to five partials there is a clear rising trend. The N in the figure stands for the average lowest partial from which the other partials were built upwards (some randomization was introduced to this parameter to secure the subjects was identifying the fundamental). This experiment thus uses higher order partials (around 10 and upwards for the upper line in Figure 6b) than the experiment that was performed for this essay and this fact would perhaps make the comparison below, between the test results in Figure 6a and Figure 6b, ineligible. However, there are several places in literature however, that describes the missing fundamental phenomenon for the lower partials as well, indicating that the partials have even more influence on pitch perception. “Low-order harmonics, particularly the third through the fifth, were found to be the most effective conveyors of missing fundamental pitch” is stated in [8]. The above experiment was however the only experiment found that had a graph that described the phenomenon in a comprehensive way.

Comparison of results between two experiments

If we compare the test results in Figure 6a and 6b we will se a rising trend in both figures as the partials increase from two to five. Even if the scales are different and the experiments are outlined differently, the trend of the results are the same. The question to be posed is: What is the reason for these trends? In the conclusions of the pitch experiment it is explained by the fact that the periodicity of a harmonic tone complex becomes better defined, particularly in a noisy background, when more harmonics are present [8]. It has earlier been stated that the periodicity detection must be using the phase information that is available from the cochlea. So the statement that “periodicity is more defined”, could only mean that more phase information is used in the periodicity-detection task. If localization of low-harmonic complexes uses
Figure 6a

A rising trend can be seen for each of the time groups as a function of the number of partials in the test tones.

Figure 6b

The picture shows the data acquired from an experiment conducted by A.J.M. Houtsma and J. Smurzynski [8]. N stands for the average lowest partial which the other partials were built upwards from (some randomization was introduced to this parameter to be sure the subjects was identifying the fundamental). In both curves an increasing trend can be seen for a correct identification of the fundamental, as a function of the number of partials (in this Figure labeled components).
phase information, as stated earlier, then a similar explanation could be given for the results in 6a. I.e. that a sound becomes easier to locate as the amount of phase information increases, as would be the case when more partials are added. The tone would be better defined, and therefore easier to localize.

One can also turn the problem around and ask what other explanations there could be for these trends to occur. In the discussion part we will return to this question at issue.

Test of significance

The next question that arises is whether the differences in the standard deviations are significant. To check this F-tests were made. For each time group three F-tests were made. This approach will increase risk of gaining a type 1 error, but this was the only method of analysis at hand. The comparisons were planned in before hand which will make the number of F-tests acceptable.

Test of hypothesis:

H1: The difference in variance between the two compared groups is large enough to be the result from two different groups.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>0.1</th>
<th></th>
<th>0.3</th>
<th></th>
<th>0.65</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-value</td>
<td>p-value</td>
<td>F-value</td>
<td>p-value</td>
<td>F-value</td>
<td>p-value</td>
</tr>
<tr>
<td>5 to 3</td>
<td>1.417</td>
<td>0.348</td>
<td>0.805</td>
<td>0.564</td>
<td>1.898</td>
<td>0.096</td>
</tr>
<tr>
<td>5 to 2</td>
<td>1.683</td>
<td>0.157</td>
<td>1.395</td>
<td>0.377</td>
<td>2.852</td>
<td>0.006</td>
</tr>
<tr>
<td>3 to 2</td>
<td>1.188</td>
<td>0.643</td>
<td>1.733</td>
<td>0.140</td>
<td>1.502</td>
<td>0.280</td>
</tr>
</tbody>
</table>

Compared groups

An criterion of p<0.05 was set to reach significance. As seen in Table 4, one of the comparisons reached significance (even at p<0.01) and three of the comparisons were close to be significant. The Null-hypothesis could therefore not be rejected in most cases, but can be rejected in the case were the 5-parital tones is compared with the 2-partial tones in the 0.65 ms. group.

Results

Only one of the F-tests in the analysis reached significance level. However, the similarity in the order of the standard deviations among the different tones between the different time groups and the rising trend of accuracy in all groups makes the relationship (the accuracy in localizing a tone and the number of partials the tone is built of) more plausible than the individual F-tests for the different time groups would show. There is a tendency that the test tones with more partials are easier to locate, but more subjects have to carry out the test to be able to reach significance for the other comparisons, which would be necessary to be able to draw reliable conclusions of the test as a whole.
Discussion

If the above relationship would have been proven statistically, what would that have meant for the resemblance between the functional mechanisms behind pitch perception and localization? How strong is the validity of the test? It would have shown that there is a resemblance in the behavior of the two mechanisms indicating that they are using the same type of information as input source. In this case the phase information from the cochlea. To prove that both percepts are dependent on phase information is of course not enough to prove that they share a common mechanism, only that they could perhaps process the information in a similar manner. If there existed a neurological preprocessing stage, whose purpose was to order and combine the phase information, before it reached the periodicity detection function or the binaural time separation function, then this preprocessing would probably use same mechanism, feeding both percepts. Many other features in nature seem to be very effectively evolved and often serve multiple purposes. But this experiment does not show if such a preprocessing stage exists. The phase information from the cochlea could maybe reach the two mechanisms directly, with no preprocessing at all. In the recently performed experiment there could also be another explanation for the results of decreasing variance as the number of partials increases. One reason could be that the tones with more partials also includes higher frequencies. The difficulty in localizing a tone could be the same as the difficulty in localizing its highest frequency. And maybe humans have a better performance in localizing higher frequencies? As a reminder, this question also has to be posed with the restriction that only phase information and not level information is available for the judgment. To avoid this possible source of error, an experiment could be outlined, where the test tones are designed by building the partials from top and downwards. Starting from the 7:th partial building down to the third. If this experiment should give the same results, then it could be established that it is the amount of partials in the tone that are the reason for the effect and not its higher frequency content. One could also look for earlier reported experiments on localization of pure frequencies and this is what the next paragraph will deal with.

In [12], W.R. Garner and M. Wertheimer performed an experiment with headphones were single pure frequencies were put to both earphones with a time delay induced to one of the channels. The scope of the experiment was to see if a subject could hear the difference in location when the leading channel was switched (ie. which channel was delayed) and how this ability depended on the frequency used. The authors concluded that it was harder to hear the difference as the frequency got higher to about 2 kHz were no differences could be heard. One could say that phase information gets poorer as the frequency rises. From this result it would be tempting to draw the conclusion that the higher frequencies could not be the explanation behind the smaller variance in the test tones with higher order partials. However, it could ironically be the opposite. When phase information gets poorer, the sensation of a pure frequency tends to be perceived from the center between the earphones, irrespectively of which time delay is used to delay one of the channels (the authors own experience from the preliminary study). The subjects perceived location of these tones could therefore be more collimated and therefore give a more narrow spread in the data, even if this location is not representative of the “true location” that corresponds to the time delay. (True location in this case refers to how a time delay between the ears usually is associated with the incoming angle of a sound which is well described in [3]). If this would be
the explanation for the smaller variance in the test results in the recently performed experiment, it would at the same time also affect the median to approach the center (the 0 on the x-axis in Fig. 5) as the higher frequencies are added. However, even if not statistically tested no such tendency is seen in the data. Therefore the author believes that the higher frequencies not are the cause for the above described effect.

It’s probably not possible to prove by listening tests alone that a preprocessing mechanism mentioned above exists. This experiment along with other experiments, like the one conducted by Nordmark, has to be complemented by further physiological studies. Not until we know more about the neurological structures belonging to the hearing sense, can we draw conclusions about the issue whether pitch perception and localization share some common functional mechanism.

Even if the statistical evidence for a positive relationship between the two percepts is not sufficient, the results show that it would be even more unlikely, that the relationship could be a negative one. And it could also be stated that the hypothesis that a tone with more partials is easier to locate can’t be rejected.

Reflection on method

The reliability of the test was enhanced by letting the subjects judge each test tone two times. In some cases outliers occurred in the data (see chapter above). These outliers reduce the reliability of the test, and if improvements were to be made to the test, it would be to reduce the possibility of getting these outliers. One way of doing this could be to introduce a training pass before the actual test. The instructions given to the subjects could also perhaps be revised, and the problem of using the extreme settings of the pointer be explained to the subjects. Another way could be to implement limits of the possible values that could be chosen. However, this approach could give some other problems to deal with when coming to the statistical analysis (like a skewed data formation for example[9]).

The difficulty in reaching significance in this test could perhaps be that the effect that is looked for is masked by a great variance among the subjects regarding the skill in localizing the tones. To go further with this investigation, one approach could be to choose maybe four subjects that have results that match the locations of the constructed test tones the most, i.e. sort out the most skillful subjects. By letting them repeat the experiment many times, a statistical analysis could be made on an individual basis; the subjects individual differences of their accuracy in localizing the different tones. This approach would reduce the influence of different skills among the subjects and the effect that is looked for would be seen more clearly.

Suggestions for further work

First of all the experiment that was performed during the work with this essay could be increased to include more subjects to be able to get the results statistically significant. A version of the experiment, that was previously mentioned, could also be outlined were fewer subjects are used but more data is acquired for each subject.
To eliminate the possible explanation that the frequency of the highest partial is responsible for the effect, the following version of the experiment is proposed. The experiment will be the same as the one performed for this essay with the difference that all partials in the tones are built from the 7:th partial and downwards. If this test should show the same tendency in the accuracy in localizing the tone depending on the number of partials, the possible explanation above would be ruled out.

It would also be interesting to do an experiment where the least discriminable shift in location is measured for the same tones that were included in this experiment. This approach would resemble other experiments where the least discriminable pitch is examined, and the result could be compared to look for similar tendencies.

When looking into references regarding the main question at issue for this essay, it seems like more experiments have been conducted on our ability to detect pitch in different stimuli than on localizing different stimuli. I think there are many experiments that examine pitch perception that could be adapted to measure localization instead. By doing this I think more similarities in behavior between the percepts would be found, that together would enlighten our knowledge on how our hearing sense works.

Further reading
References [2], [5] and [6] has not been directly referenced in this essay. They have though served as a source of inspiration to the author and may come in use for further work in this area. For the interested reader will here follow a short summary of those references.

[2] This paper describes one relationship between the two percepts, localization and pitch perception. Here the answer to the issue mentioned in the introduction, whether a higher pitch tends to be perceived as coming from a higher vertical place, will be found. The answer has nothing to do with phase information however, and therefore it was not referenced in this essay.

[5] This paper describes an experiment where low frequency temporal information is presented to locations in the cochlea tuned to high frequencies. The experiment shows that tonotopic representation (ie. which nerves in the cochlea submits the information) is crucial to complex pitch perception.

[6] This essay by G. Oster gives a very good summary of different experiments and theories concerning auditory beats in the brain. Amongst other effects, the difference between “monaural beats” (the combination of two slightly different frequencies is fed to both ears, or only one ear is used) are compared with “binaural beats” (two slightly different frequencies are presented to different ears). From this some conclusions may be drawn about how our brain processes auditory information.

Acknowledgements
This work was supervised by doc. Pehr Sällström, whom I want to thank for being a good source of inspiration and for the valuable help with the essay.
References


