Apple Hearing Demo Reel

Apple Computer Technical Report #25
Malcolm Slaney and Richard F. Lyon
malcolm@apple.com, lyon@apple.com

Speech and Hearing Project
Advanced Technology Group
Apple Computer, Inc
Cupertino, CA 95014

1. Introduction

This document accompanies the video tape “Apple Hearing Demo Reel” and describes each of the clips on the tape. The video is not meant to stand alone as a representation of our work. Instead it complements our papers and talks by showing material that isn’t easily published in journals. We hope that researchers who are already familiar with our work will benefit from this video.

In most cases the video tape and this document do not represent the final word, but instead show the current state of our research efforts. They have been assembled so that other researchers interested in hearing can see the animations that we have found so revealing. When possible, we provide references to our work and the work of others that can provide more information about each clip.

Most of the examples on the tape show correlograms. A correlogram is a two dimensional representation of sound that we find to be very rich in information. The correlogram shows a picture of sound where spectral frequency is along the vertical axis and short–time correlations are shown on the horizontal axis. This representation is described by Licklider [Licklider51] and is described in Clips 3 and 7.

We are especially struck by the similarity of the auditory perception and our visual perception of the correlograms. There is neurophysiological evidence that sounds are represented in the brain as two or three dimensional maps. There is little evidence that a specifically correlogram–like image is computed by the brain, but the examples shown on this tape are compelling enough to pique our interest. We are currently exploring the use of the correlogram for modeling the performance of human listeners in pitch perception and sound separation experiments.

The clips on this tape have been created during 1989 and 1990. The correlograms were computed using our ear model [Slaney88] followed by an implementation of the correlogram model running on a Cray supercomputer at Apple Computer. Sounds were digitized, stored on computer, and then written onto video tape. The correlograms were then computed and each frame was individually written onto tape. While the computations are fairly fast (approximately 2 seconds of Cray CPU time for each 1 second of audio) the VCR can write only one frame every 10 seconds. We put a beep at the start of each audio selection so we can synchronize the audio and video on the tape. Once the start of the sound is found on the tape the rest of the procedure is automated.

We appreciate the help that we’ve received from Richard Duda, Bill Stafford, Ray Meddis, Roger Reynolds, Scott Stein, and Kent Koeninger in preparing this tape. Many of our ideas were shaped by discussions with people at the Hearing Seminar at Stanford’s CCRMA and at Caltech's computational and neural systems group. Also we appreciate receiving permission from the Acoustical Society of America to publish some of the sounds from their *Auditory Demonstrations* Compact Disk [Houtsma87].

2. Cochlear Motion (Starts at 3:02, length is 2:00)

The first three animations show the pressure wave along the basilar membrane as tones propagate through the cochlea. These animations were computed using *Mathematica* [Wolfram88] and the simulations are based on a model of the hydrodynamics of the cochlea as developed by Richard Lyon and Carver Mead [Lyon89]. The goal of these demonstrations is to show how the basilar membrane propagates waves and distributes the spectral energy of a sound along its length. Each of the animations is accompanied by an audio commentary describing the features of the animations.

2.1 Cochlear Motion—Single Tone (Starts at 3:02, length is 0:35)

This simulation shows how a pressure wave in the cochlea travels at first with a rather low amplitude and a high velocity. As the wave travels it slows down as its velocity decreases and thus its amplitude grows. Eventually the wave is travelling so slowly that it dies down due to membrane losses.

Audio Commentary on Tape:

> "This *Mathematica* animation shows the pressure wave at the membrane of a pure tone as it propagates down the cochlea. Where it first enters the cochlea, on the left, it has a high propagation speed. As the wave travels, it slows down and the energy starts to bunch up. Beyond a point of maximum response the pressure wave quickly dissipates. The losses in the cochlea are still relatively small here, but the wave is traveling so slowly that it is strongly attenuated before it travels much farther down the cochlea. This very slowly traveling wave accounts for the steep attenuation measured by hearing researchers. A rather broad 'pseudoresonance' describes the resulting response."

2.2 Cochlear Motion—Two Tones (Starts at 3:38, length is 0:25)

This animation shows two tones propagating down the cochlea. The first tone, at three times the frequency of the second tone, peaks first. Note how the peaks in the left wavelet are going up and down at a much faster rate than those on the right. This shows that different frequency tones are resolved by different parts of the cochlea.

Audio Commentary on Tape:

> "It is important to realize that this pseudoresonance occurs at different locations for different frequencies. This animation shows the pressure due to two tones at different frequencies. The higher frequency tone, at three times the frequency of the low–frequency tone, peaks considerably earlier. With the scaling assumptions we have used, the wavelength at the response peak is independent of frequency."

2.3 Cochlear Motion—Gray Single Tone (Starts at 4:03, length is 0:59)

This animation shows a gray–scale representation of the pressure in the cochlea. The previous two animations showed the pressure along the basilar membrane as a sound propagated down its length. This animations now looks at the pressure in the fluid. As the pressure wave travels down the Basilar Membrane it exerts pressure on the fluid. The pressure field is shown in this animation for a single audio input tone. The basilar membrane is shown along the center line.

Audio Commentary on Tape:

> "The change in propagation speed means that the long–wave, or one–dimensional, approximation is only valid in the early part of the wave's travel. Near the response peak the wavelength is short enough that energy can flow both down the cochlea and perpendicular to the membrane, so a two–dimensional model is needed. This *Mathematica* animation shows the pressure wave due to a pure tone as it travels down the cochlea. Darker and lighter regions correspond to pressures above and below the average.
>
> "On the left, where the propagation speed is high, the wave in each chamber is essentially one dimensional and the wave travels entirely along the direction of the basilar membrane. As the wave slows down the energy starts to bunch up and there is now a variation in pressure in the direction perpendicular to the length of the cochlea.
>
> "After a short distance the energy in the wave is completely dissipated. On the right, very faint waves are actually traveling at an angle, carrying their energy into the basilar membrane where the energy is dissipated in membrane losses."

3. Correlogram Narration Original (Starts at 5:12, length is 1:30)

This is our original description of a correlogram [Licklider51, Lyon84]. It explains the details of the correlogram in a slightly more technical form than the Leonardo clip that follows (Clip 7). Unfortunately some of the vowels are a bit weak and the singing isn't nearly as good as the Leonardo example. This animation does explain a few details of the correlogram that are unstated by Clip 7.

Audio Commentary on Tape:

> "You are observing a correlogram of the sounds you are now hearing. This animation was computed on the Cray X/MP at Apple. We believe this is the first time a real time display of the correlogram has been shown. While the correlogram is still a theory, we believe the animation you are watching is a plausible representation of the activity of the neurons in your brain as you listen to this narrative.
>
> "The correlogram is important because it shows several relevant aspects of a speech signal. The dark horizontal lines indicate high energy in a particular frequency band, or formants, while the vertical lines represent pitch information.
>
> "Only voiced sounds have a pitch. During unvoiced sounds, like the letter "s", the vocal cords are relaxed and the sound is more noise like. Listen to the sound and watch the
>
> "SSSSSSSSAAAAAAAAASSSSSSSSSYYYYYYY
>
> "The correlogram does a good job of separating the formants and the pitch information in the signal and integrating them into a unified representation. Watch how the correlogram changes as I change the pitch of this vowel
>
> "/i/ eee /i/ eee (as in beet).
>
> "If the pitch is held constant and the vowel changes then only the formants move. For example:
>
> "/i/ eee /u/ ooo /R/ ir"
> (beet) (boot) (bird).

4. Reynolds–McAdam's Oboe (Starts at 7:00, length is 1:04)

A phase vocoder was used to split the sound from a single oboe into two parts. The even harmonics are grouped separately from the odd harmonics. Each set of harmonics is modified by independently jittering the pitch; the result is that a single oboe splits into two different sounds.

This clip is a very good example of auditory streaming. The sound was created by Roger Reynolds (now at UCSD) for a composition he wrote called "Archipeligo" [Reynolds83] and is used here with permission. The idea for this example came from Steven McAdams, whose thesis [McAdams84] talks about what needs to be done to a sound to make it sound like more than one auditory object [Bregman90].

We are excited about this example for two reasons. First, it is a very good example of auditory streaming. A single sound, an oboe in this case, is first heard as a single instrument and then splits into two sounds, a clarinet and a soprano. It is a very good demonstration of what it takes for sounds to separate and to sound isolated. The second reason we are excited about this clip is that it shows the power of the correlogram. We think the correlogram is a very rich representation and clearly shows the two sounds as distinct visual objects.

The cynics in our audiences have pointed out that we have just translated one impossible problem (auditory grouping) into another impossible problem (visual understanding). This is correct but we feel that if we can see the auditory objects then it might be possible to design algorithms that can do the same. This is a current research direction and one possible approach is described in Clip 5.

In this demonstration, the pitch of the two sets of harmonics is at first fixed. This is the original oboe sound. Then, after a few seconds the pitches of each set of harmonics are independently jittered and the sound separates into two components. The odd harmonics sound like a clarinet since a clarinet has mostly odd harmonics. The even harmonics are now an octave higher (since the fundamental is now twice as high as before) and sound like a soprano. An important reason that the harmonics sound like a realistic clarinet and soprano is that the vibrato rate (5–6Hz) is very natural.

There are four pieces to this example (each approximately 16 seconds long):
1) Combined sound (first fixed and then jittered)
2) Just the even harmonics (soprano)
3) Just the odd harmonics (clarinet)
4) Combined again.

5. Motion (Starts at 8:308, length is 1:19)

One principle that seems to be important in separating sounds into separate auditory objects is the idea of common fate. Sounds can be grouped by common amplitude and frequency fluctuations. Thus parts of an acoustic signal that all turn on at the same time or that are changing pitch synchronously tend to be fused into a single auditory object. (See also Clip 11.)

We are investigating the use of an algorithm known as optic flow to measure motion in a correlogram. The optic flow is computed by looking at the derivatives of the images with respect to x, y and time. These derivatives give some information about the motion of objects between the two frames of a video sequence. See [Horn86] for more information about optic flow.

But velocity is a two–dimensional quantity, and the velocity constraint provides only one equation. A second equation is needed to regularize this problem, and generally some sort of smoothness constraint is added. In this case we try to find a solution that minimizes the curvature of the resulting velocity fields. These equations are solved by doing a Jacobi iteration looking for a fixed point solution.

5.1 Motion—Test (Starts at 8:08, length is 0:02)

This is a simple example that is used to test our motion algorithms. In these animations we have used brightness to encode the energy in the image (white is high energy) and we have used color to encode the direction of movement. When an energy in the picture is moving to the right it will be shown in bright red. When it is stationary it is white and when it is moving to the left it will be shown as bright green. We apologize to men in the audience who are red–green color blind. When we first chose the colors for these examples we weren't thinking about that. When we redo these animations we will choose less ambiguous colors.

5.2 Motion—McAdam's Oboe (Starts at 8:17, length is 1:09)

This demonstration shows a correlogram of the Reynolds/McAdams oboe coded so that motion to the left is colored green and motion to the right is red. Brightness corresponds to energy in the correlogram.

This is work in progress. The code used to generate this animation is not working quite right. (We had convergence problems when we made this video and the upper frequencies are quite noisy) We believe that this animations shows the technique can work well enough so that it is worth our while to pursue.

We have also transformed the correlogram so that now the horizontal axis (short–time autocorrelation time delay) is logarithmic. In a standard correlogram (with linear time) a change in pitch means that the image stretches along the horizontal axis. By using a logarithmic time axis a pitch change becomes a rigid shift.

This example shows that it is possible to code pieces of a correlogram by velocity. Eventually we hope to segregate portions of the correlogram based on their common velocity (rising or falling pitch and formants) and then to isolate individual auditory objects. It remains to be seen whether this will work or will be an accurate model of how human listeners perform auditory scene analysis, but this is a good first step.

Note that this is a very preliminary result using code that isn't working quite right. The sense of motion is best seen if you can run the animation around 12 seconds in very slow motion (no more than a couple of frames per second).

6. Running to You (Starts at 9:31, length is 3:27)

This animation shows the correlogram of a musical piece and its components. Separating the voice from such a mixture is the Holy Grail of the hearing effort at Apple. Human listeners can certainly understand the words that accompany this music; why can't machines? If we ever get to the point where we can separate out the voice from this background "noise" then a typical office environment should be easy.

This video clip shows three correlograms. The original song was recorded onto multi–track tape and then mixed so that the vocals were on one channel and the instrumentals were on the other channel. We digitized the two channels independently and added them together to make the sound track on this tape. Correlograms were then individually computed of the vocals, the instrumentals, and the mix.

The middle frame of this animation shows the correlogram of the sound you are actually listening to. The top frame shows the vocal track while the bottom frame shows the instrumental track. (Note that there is some bleeding of the sound from the instrumental track to the vocal track.)

It is unclear whether the correlogram is up to this task. We can certainly see evidence of the singer's pitch in the middle correlogram. This task is made harder because the pitch of the woman's voice harmonizes with the music. This means that her pitch will often land on top of the musical components and thus be hidden. More work remains to be done…

This piece, called "Running to You" was composed and performed by SSK (Steve Milne, Steve Squyres, and Kalena O'Malley).

7. Correlogram Narration—Leonardo (Starts at 13:03, length is 0:58)

This is an introductory description of a correlogram. It was originally done for an internal Apple presentation where Leonardo Da Vinci explored various aspects of Apple technology. It explains the various features of a correlogram and shows how the correlogram changes as the pitch and formants move. The speaker/singer here is Peter Cavanaugh. It is similar to Clip 2 but has fewer technical details and better singing.

Audio Commentary on Tape:

> "Sight is the lord of astronomy; the prince of Mathematics. It counsels and corrects all the arts of mankind.
>
> "What fun! Here, observe! Ahhh Ahhh Ahhh (with rising pitch). See? The dark vertical lines follow what the muscles in my throat do.
>
> "Now, I will change the shape of my mouth but keep the same pitch with my throat. Ahhh, Eeee, Iyyy, Ohhh, Oooh. Yes! Yes! The horizontal lines determine what sounds come, that come out of my mouth.
>
> "Fascinating! If I had such a machine I could tell who it is that is speaking by the vertical pitch (sic, should be lines) and what was said by the horizontal lines."

8. Reading Machine (Starts at 14:09, length is 1:37)

This clip, one segment from Apple's video tape "Project 2000", shows a very compelling application of speech recognition. In this "vision" of the future an adult is tutored by a "Reading Machine." The machine can understand his speech as the words on the screen are read. When the reader is unsure of a word he can ask for the pronunciation.

This is an application of speech recognition that is doable today. In general the reader already knows how to pronounce the words and is a motivated learner. The perplexity, a measure of the complexity of the grammar, is low and the speech rate is rather slow.

9. Pitch (Starts at 15:53, length is 3:21)

During 1989 we did some work building a pitch detector that mimics the human perceptual system [Slaney90]. Traditional approaches base a pitch decision on features of a relatively primitive representation such as the waveform or spectrum. Our pitch detector uses an auditory model. Unlike the simpler techniques, our perceptual technique works for a wide range of pitch effects, and is robust against a wide range of distortions. Meddis has compared this type of approach with human performance [Meddis91].

The technique used was first proposed by Licklider [Licklider51] as a model of pitch perception, but it has not been taken seriously as a computational approach to pitch detection due to its high computational cost.

The representation used by our pitch detector, which corresponds to the output of Licklider's duplex theory, is the correlogram. This representation is unique in its richness, as it shows the spectral content and time structure of a sound on independent axes of an animated display. A pitch detection algorithm analyzes the information in the correlogram and chooses a single best pitch.

There are many signals, such as inharmonic tones or tones in noise, that do not have a periodic time or frequency–domain structure, yet humans can assign pitches to them. The perceptual pitch detector can handle these difficult cases and is thus more robust when dealing with the common cases. We expect that future systems will benefit by using this approach, or a cost–reduced version.

There is still considerable freedom to devise algorithms to reduce the rich correlogram representation to a pitch decision. The results we reported are from a relatively simple algorithm, which does not address many of the subtle issues involved in a pitch tracker for use in a real system. Our algorithm picks a pitch for each frame of the correlogram, and does not address the decision of whether there is a valid pitch (as in the voiced/unvoiced decision in speech processing). Nor does it attempt to enforce or utilize frame–to–frame continuity of pitch. Humans have complex strategies for making such decisions, depending on the task (for example in music, jumps in pitch define the melody, so continuity should not be enforced).

More work is needed to tune this model to accurately predict the pitch of inharmonic sounds. The results so far are consistent with the so–called "first effect", but not with the more subtle "second effect." [Small70]

The clips that follow show two "correlograms" and two summaries synchronized to the audio. The top correlogram is a normal correlogram. The second correlogram has been enhanced with a simple local operator that enhances the vertical (pitch like) structures in the correlogram. The energy in the correlogram is then summed along vertical columns to create what Ray Meddis calls a summary correlogram. This graph represents the likelihood of a pitch at each time delay. Peaks in this correlogram summary correspond to time scales in the cochleagram where many channels have a periodicity. Finally the last frame shows our single best guess for the pitch.

Following each of the pitch examples on this tape there is a summary of the perceived pitch as a function of time.

9.1 Pitch—Leonardo Vowels (Starts at 15:53, length is 0:06)

This animation shows our perceptual pitch detector working on the singing vowels from the Leonardo demonstration.

Note that during the second portion of this demonstration there are often octave errors. This pitch detector does not take into account the fact that pitch often changes smoothly. Thus in the perceptual pitch detector the pitch of each frame is independent, while human listeners generally assume that the pitch varies smoothly.

9.2 Pitch—Westminster Chimes (Starts at 16:10, length is 0:16)

This demonstration shows that the perceptual pitch detector can recognize the pitch even when there is a high level of noise. In this example, part of Demonstration 22 on [Houtsma87], low pass noise is added to a melody that is played with alternating spectral and virtual pitch. The low pass noise completely masks the spectral pitch but the virtual pitch can still be heard and is found by the perceptual pitch detector. The signal to noise ratio is approximately –20dB.

This is the description of clip:

> "This demonstration uses masking noise of high and low frequency [only low is used in this video clip] to mask out, alternately, a melody carried by single pure tones of low frequency and the same melody resulting from virtual pitch from groups of three tones of high frequency (4th, 5th and 6th harmonics). The inability of the low–frequency noise to mask the virtual pitch in the same range points out the inadequacy of the place theory of pitch."

9.3 Pitch—Shepard Tones (Starts at 16:33, length is 0:32)

Pitch can be ambiguous. This example from Example 27 on the ASA *Auditory Demonstrations* CD [Houtsma87] shows the auditory equivalent of a rotating barber pole. In this case the pitch of this signal always appears to decrease.

Human listeners tend to hear a single pitch. The pitch is heard to go down until the listener loses his/her concentration or the pitch has become so low as to be nonsensical. Then a new pitch is chosen by selecting a new octave and the process continues.

In our perceptual pitch detector there is no temporal continuity. During each frame of the correlogram the most likely pitch is chosen. That is why you often see the estimated pitch oscillate between two choices at times when two pitches are likely. This also shows up in the final pitch summary as two pitches for some of the times.

Here is the description of the clip. We only show the correlogram and pitch of the continuous Shepard tones in this tape:

> "One of the most widely used auditory illusions is Shepard's (1964) demonstration of pitch circularity, which has come to be known as the "Shepard Scale" demonstration. The demonstration uses a cyclic set of complex tones, each composed of 10 partials separated by octave intervals. The tones are cosinusoidally filtered to produce the sound level distribution shown below, and the frequencies of the partials are shifted upward in steps corresponding to a musical semitone (Δ 6%). The result is an "ever–ascending" scale, which is a sort of auditory analog to the ever–ascending staircase visual illusion."

9.4 Pitch—First Effect of Pitch Shift (Starts at 17:11, length is 0:22)

Consider a three–tone complex produced by AM modulation of a 1kHz tone at 100 Hz. This will be perceived with a pitch of 100Hz since the three tones represent the 9th, 10th and 11th harmonics of a 100 Hz fundamental. Now, if the carrier frequency is moved up 20Hz to 1020 Hz the tones no longer form a harmonic complex with a 100 Hz fundamental. The sound is harmonic with a fundamental of 20Hz but frequencies this low are not usually heard as a pitch. Instead human listeners perceive the sound to have a pitch of 102 Hz, as if the sound is still harmonic and is approximately the 9th, 10th and 11th harmonics of a 102 Hz fundamental. This is known as the first effect of inharmonic pitch shift [Schouten62]

An easy way to generate such a signal is to amplitude modulate a carrier frequency. This can be written

$$\sin(F_C)\ [1 + \sin(F_m)].$$

where F_C is the carrier frequency and F_m is the modulation frequency. This expression can be expanded into its Fourier components and has the following components

$$1/2\ \cos(F_C - F_m) + \sin(F_C) - 1/2\ \cos(F_C + F_m).$$

or components at the carrier frequency and the carrier frequency plus and minus the modulation frequency.

This clip shows the perceived pitch of a pure tone that has been amplitude modulated with a 100Hz carrier (100% modulation). The carrier frequency increases slowly from 600Hz to 1200Hz. At each multiple of 100Hz the signal is harmonic and between these points the pitch is ambiguous. The perceptual pitch detector chooses the most likely response but often oscillates between two different pitches.

In general, the pitch of this signal varies directly with the pitch of the carrier. Thus at 900Hz the perceived pitch is 100Hz since the signal is harmonic. If the carrier frequency is raised by Δ then the pitch goes up by a factor of Δ/9. This can be seen by examining the fine time structure in the modulated waveform.

9.5 Pitch—Second Effect of Pitch Shift (Starts at 17:39, length 0:12)

Researchers noticed that the pitch of inharmonic signals did not vary exactly as expected [Schouten62]. This demonstration shows the perceptual pitch detector's analysis of amplitude modulating a carrier that varies between 1200 and 2400 Hz with a modulating tone of 200Hz (90% modulation). It does not show the predicted extra shift. Some people think that the second effect of pitch shift is caused by non–linear

distortions in the cochlea. Our current implementation of the correlogram uses a linear cochlear filter so there are no combination tones generated.

9.6 Pitch—Spectral vs. Virtual Demo (Starts at 18:52, length is 0:15)

This demonstration aims to illustrate the difference between spectral and virtual pitch. The time domain waveform, the spectrum and the perceived pitch of a harmonic signal are shown. The first half (8 seconds) of this clip demonstrate the sound as harmonics are added (first, just a 440 Hz cosine wave, followed by 440 Hz and its second harmonic, and so on). This half shows how the timbre changes (becomes "richer") as harmonics are added.

The second half of the demonstration shows the effect as the lower order harmonics are removed. First the fundamental is removed and then the second harmonic and then the others until only the highest harmonic (at 3520 Hz) remains. This shows that while the timbre changes, the pitch remains the same until there are only two harmonics.

Unfortunately, the audio level was set a bit high on this demo so there is a lot of distortion. Sorry about that. Someday we'll have to redo it. Sounds similar to this demonstration can be heard on the "Duda Tones" example later on the tape.

9.7 Pitch—Ambiguous Sentence (Starts at 19:13, length is 0:07)

This example shows our pitch perception algorithm working on a sentence with only voiced sounds [Slaney90]. This example was provided by Patti Price at SRI and illustrates two ways that the sentence "They may wear down the road" can be said.

10. Duda Tones (Starts at 19:27, length is 1:10)

This animation was produced in conjunction with Richard Duda of the Department of Electrical Engineering at San Jose State University during the Summer of 1989. Thanks to Richard Duda for both the audio examples and the explanation that follows [Duda90] The following demonstrations are shown in this clip:

1) 200Hz tone

2) The first eight harmonics of a 200Hz tone.

3) The first eight harmonics are added one at a time (two per second)

4) Same as 3 but faster

5) Now the sequence is reversed. The first eight harmonics are present and then removed one at a time from the top.

6) Same as 5 but now the harmonics are removed from the bottom

7) All eight harmonics but now add vibrato to each one to make it separate out.

A finite Fourier series is a classical representation of a single periodic waveform as a mixture of component pure tones. When one listens to such synchronized mixtures, one normally hears a single, coherent sound rather than the separate harmonic components or "partials." However, there are interesting circumstances under which the individual harmonics can be heard very clearly.

One such circumstance is when the composite tone is built up or broken down sequentially. Another is when the harmonics are individually modulated, whether in amplitude, frequency or phase. A third is when there are only a few harmonics and they form a familiar musical chord. In these situations, one perceives the sound to split into separate and distinct "voices" or "streams" that can emerge from the composite tone.

Several experiments were performed to see how amplitude and frequency modulation of the harmonics affect both auditory perception and the correlograms. In the cases described below, the signals were sawtooth waves, the first eight harmonics of a 200–Hz fundamental. The following observations were made:

1. If the entire signal is presented abruptly, it sounds like a single, somewhat buzzy sound source, the component harmonics not being noticeable without concentrated effort. If the signal is built up sequentially by starting with the fundamental and rapidly bringing in the harmonics sequentially

(less than 50–msec between entrances), the result still sounds like a single source but with changing tone color during the "attack." However, if the harmonics are brought in abruptly but slowly (say, 500–msec between entrances), each new harmonic sounds briefly like a new sound source. The highest harmonic tends to remain separated from the rest for a few seconds, but it eventually fuses with the lower harmonics into a single source.

2. If the harmonics are abruptly turned off one at a time, the psychological effect is quite different. It is perhaps best described as sounding like a single sound source whose tone quality or "timbre" undergoes abrupt but subtle changes. This is an example of a rather general phenomenon, in which sudden decreases in amplitude (offsets) are generally much less salient than sudden increases in amplitude (onsets).

3. If the frequency of one of the harmonics is modulated a small amount at a sub–audio rate, its sound vividly emerges as a separate "voice." A one–percent sinusoidal modulation of the frequency of any of the first 12 harmonics of a 200–Hz sawtooth, for example, produces this effect clearly, as does a one–percent step change in frequency. The effect is strong and striking, indicating that the auditory system is highly sensitive to changes in frequency. (We do not know if this is because there are cells that are directly sensitive to frequency modulation, if cells with narrow spectral tuning are seeing sound come in and out of their sensitive region, or if the change in the time–domain wave shape is directly perceived.)

The existence of the separate components is also revealed in the correlograms when they are recorded on videotape and viewed dynamically in real time. Components having separate onsets or offsets are clearly visible when they jointly appear or disappear. Similarly, components having common frequency modulation stand out through their joint motion. Although no calibrated psychoacoustic measurements were made, the brief time required for an amplitude or frequency change to be seen in the correlogram seemed comparable with the time required to hear a new voice emerge. However, when the modulation was stopped, the dynamic response of the correlogram seemed much faster than the time required to hear separate components merge back into one sound stream.

These results imply that the information needed to separate harmonic components is present in the correlogram. Furthermore, the primitive nature of the signals implies that the source formation and separation mechanisms do not depend on high–level domain knowledge, but can be performed in the early stages of processing and are properly part of a model of early audition. While the adaptivity of perceptual mechanisms precludes simple explanations, the importance of common modulation implies the need for comodulation detection and grouping functions in any model of the early auditory system [Mont–Reynaud89].

11. Duda Vowels (Starts at 20:47, length is 2:16)

This animation was produced in conjunction with Richard Duda of the Department of Electrical Engineering at San Jose State University during the Summer of 1989. Thanks to Richard Duda for both the audio examples and the explanation that follows [Duda90].

The following correlograms are shown here:

1) /a/ vowel (140–Hz fundamental, 5% 6–Hz vibrato)
2) /i/ vowel (132–Hz fundamental, 5% 6–Hz vibrato)
3) /u/ vowel (125–Hz fundamental, 5% 6–Hz vibrato)
4) All three vowels with no variation (world's most boring cocktail party)
5) 5% FM—Three vowels with sequential 5% pitch vibrato
6) 1% FM—Same as 5 but only 1% (barely audible)
7) 100% AM—Three vowels with sequential 100% amplitude modulation
8) 25% AM—Same as 7 but only 25% (barely audible)
9) 6dB Increase—Three vowels with sequential +6dB step increases in amplitude
10) 6dB Decrease—Same as 9 but now a –6dB step decrease
11) 5% Pitch Increase—Three vowels with sequential 5% increase in pitch
12) 1% Pitch Increase—Same as 11 but now with sequential 1% increase in pitch
13) 5% Pitch Decrease—Three vowels with sequential 5% decrease in pitch
14) 1% Pitch Decrease—Same as 13 but now with sequential 1% decrease in pitch

Vowel sounds can be thought of as periodic signals with very special spectral patterns. Their spectral envelopes are dominated by the formant resonances, which are basically independent of the fundamental frequency. Through learning spoken language, people become particularly sensitive to the spectral envelopes for the vowels in that language.

In his doctoral thesis, Stephen McAdams showed that when different vowels having the same fundamental frequency were mixed, the resulting mixture did not have a vowel–like quality; however, when the glottal pulse train for any one vowel was frequency modulated, the sound of that vowel would "emerge" as a separate, recognizable sound stream [McAdams84]. The effect was very strong, again suggesting that the auditory system makes central use of the comodulation of harmonic components to separate sound sources.

Several experiments were performed to see how both amplitude and frequency modulation of the glottal pulse train affected the auditory perception and the correlograms for the vowel mixtures. All experiments used the same three synthetic vowels: /a/ (as in "hot"), /i/ (as in "beet") and /u/ (as in "boot"). These vowels were synthesized using a cascade model due to [Rabiner68]. Specifically, a pulse train with fundamental frequency f0 was passed through a cascade of two single–pole glottal–pulse filters, three two–pole formant filters, and a first–difference stage to simulate the effects of radiation. The glottal–pulse filter has poles at 250 Hz, and the formant resonances had a 50–Hz bandwidth. The following table lists the fundamental and formant frequencies:

	f0	f1	f2	f3
/a/	140	730	1090	2440
/i/	132	270	2290	3010
/u/	125	300	870	2240

The correlograms of these individual vowel sounds and their mixture are shown on the video tape. One notes at once the distinctly different visual patterns of these three vowels. The horizontal organization (rows of energy blobs) reveal the formants, the high frequency formants for the /i/ being particularly distinctive. The vertical organization (columns of energy blobs) reveal the distinctly different pitch periods, which are about a semi–tone apart. Note, however, that there is no clear pitch period in the mixture, whose low–frequency organization is murky. The mixture sounds comparably murky; like a dissonant blend of unidentifiable tones.

A variety of different synthesized vowel mixtures were produced by modulating the glottal pulse trains in different ways. The standard method was to create 4–second pulse trains for the three vowels as follows:

Vowel	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0 (sec)
/a/			MMM						
/b/				MMM					
/c/					MMM				

For example, the pulse train for /a/ was held steady for 1 second, modulated (M) for the next 0.5 second, and then held steady again for the remaining 2.5 seconds. Signals with the following kinds of modulation were generated in this fashion (not all are shown on the tape):

1) Sinusoidal 6–Hz frequency modulation: 0.2%, 0.5%, 1%, 2% and 5% percentage modulation.

2) Sinusoidal 6–Hz amplitude modulation: 5%, 25%, and 100%.

3) Step amplitude change: –6dB, –3dB, +3dB, +6dB. In these experiments, the amplitude was changed for 0.5 seconds and then restored to its initial value.

4) Step frequency shift: .1%, .2%, .5%, 2%, 5%. In these experiments, once the frequency was shifted, it was held steady, rather than returning to its original value.

The perceptual character of these signals can be summarized as follows:

1) The steady vowel mixture sounds like an uninteresting, dissonant, buzzy chord with no vowel qualities. With 5% frequency modulation (vibrato), the vowels clearly and effortlessly emerge from this mixture. At 1% modulation the effect is still clear, but 0.5% is marginal. When one knows what to listen for, changes can be heard (or at least imagined) with .2% and even .1% modulation.

2) Frequency shift is even more effective than sinusoidal modulation. The vowels “come in” much like the harmonics do when a periodic wave is built sequentially. Furthermore, the ear tends to hold on to the higher–pitched sounds, so that the /i/, with its prominent high–frequency formants, clearly persists as a separate sound stream even after all signals are again steady.

3) When the frequency shift is 5% or 1%, one has a clear sense of both the direction and amount of pitch change. With the .5%, .2% and .1% shifts, one is aware that something has changed, but the pitch seems the same.

4) Sinusoidal amplitude modulation (tremolo) is not particularly effective in separating sounds. Although 5% and 25% modulation patterns were certainly noticeable for the vowels in isolation, they produced inaudible to marginal changes in the mixture. At 100% modulation the vowels “emerge,” but not as well as they do with 5% frequency modulation.

5) A 6–dB step change of amplitude (100% up, 50% down) is clearly audible, including the offset when the last vowel returns to its original level. A 6–dB increase causes the vowel to stand out. However, a 6–dB decrease creates an unsettling awareness of change, but with no well–defined vowel sound. In fact, one frequently hears the vowel only when its amplitude is restored to its original value. The effect is still obtained with a 3–dB change (40% up, 30% down), but it is beginning to be marginal.

All of the changes that could be easily heard could also be easily seen in videotapes of the correlograms. Even the small 1% frequency shifts produced clearly visible motions. While the formants can also be seen in the correlograms, the untrained eye is not naturally sensitive to these spectral patterns. That is, recognizing the vowel from viewing the changes is not obvious, and would require an ability to read correlograms similar to the ability of trained spectrogram readers to identify vowels in spectrograms [Zue85]. However, it seems likely that a vowel–recognition procedure that worked on the correlograms of isolated vowels would also work on the fragments of correlograms separated by comodulation.

However, these experiments also revealed a phenomenon that might limit a recognition strategy based on unsophisticated motion detection procedures. Low–frequency beats between the three fundamental frequencies produced moving patterns in the correlogram even in the absence of any modulation of the pulse train. The difference between motion due to beats and motion due to comodulation or other temporal changes in the acoustic input needs to be better understood before grouping models based on comodulation can be developed.

12. Duda Pierce Tones (Starts at 23:17, length is 1:33)

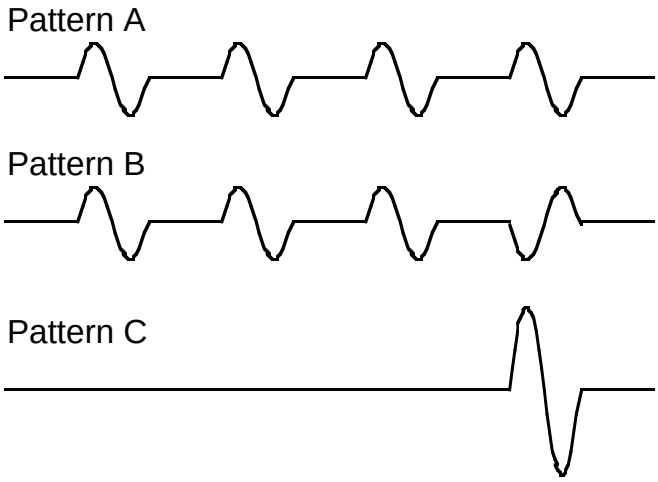
This animation was produced in conjunction with Richard Duda of the Department of Electrical Engineering at San Jose State University during the Summer of 1989. Thanks to Richard Duda for both the audio examples and the explanation that follows and to John Pierce for calling this experiment to our attention.

Researchers in psychoacoustics have long looked to cochlear models to explain the perception of musical pitch [Small70]. Many experiments have made it clear that the auditory system has more than one mechanism for pitch estimation. In one of these experiments, Flanagan and Guttman used short–duration impulse trains to investigate two different mechanisms for matching periodic sounds, one based on spectrum and one based on pulse rate [Flanagan60]. They used two different impulse trains, one having one pulse per period of the fundamental, the other having four pulses per period, every fourth pulse being negative . These signals have the interesting property that they have the same power spectrum, which seems to suggest that they should have the same pitch. The standard conclusion, however, was that below 150 pulses per second the trains “matched” if they had the same pulse rate; they “matched” on spectrum only when the fundamental frequency was above about 200 Hz.

[Pierce89] modified this experiment by replacing the pulses by tone bursts—short periods of a 4,800–Hz sine wave modulated by a raised–cosine Hamming window. In essence, he used Flanagan and Guttman’s

pulses to amplitude modulate a steady high–frequency carrier. His purpose in doing this was to narrow the spectrum, keeping the large response of the basilar membrane near one place (the 4,800–Hz place), regardless of pulse rate.

To be more specific, Pierce used the three signal "patterns" shown below. All have the same burst duration, which is one–eighth of a pattern period. Pattern *a* has four bursts in a pattern period. Pattern *b* has the same burst rate or pulse rate, but every fourth burst is inverted in phase. Thus, the fundamental frequency of *b* is a factor of four or two octaves lower than that of *a*. Pattern *c* has only one burst per pattern period, and thus has the same period as *b*; in fact, it can be shown that *b* and *c* have the same power spectrum. Thus, *a* and *b* sound alike at low pulse rates where pulse–rate is dominant, and *b* and *c* sound alike at high pulse rates where spectrum is dominant. Pierce observed that the ear matches *a* and *b* for pattern frequencies below 75 Hz, and matches *b* and *c* for pattern frequencies above 300 Hz. He found the interval between 75 and 300 Hz to be ambiguous, the *b* pattern being described as sounding inharmonic.



Pierce's tone bursts. Patterns *a* and *b* have the same pulse rate frequency, while *b* and *c* have the same power spectrum. Here the test sounds are shown with one cycle per burst.

To see if and how these results are reflected in correlograms, a similar set of tone burst signals were generated. The only difference between our signals and Pierce's signals was due to differences in the digital sampling rate used. To get a Fourier spectrum with minimum spectral splatter, Pierce imposed two requirements:

1) The tone–burst frequency f_b was set at half the Nyquist rate. Where Pierce's 19,200–Hz sampling rate led to f_b = 4,800 Hz, our 16,000–Hz sampling rate forced f_b down to 4,000 Hz.

2) Each burst had to contain an exact integral number n of cycles. This number, n, is a major parameter for the experiments, ranging from 1 to 128. If the pattern period is T, then to obtain exactly n cycles of frequency f_b in time T/8 requires that f_b T/8 = n, so that $T = 8n/f_b$.

Thus, to obtain the same spectral characteristics, we had to use different numerical values for the tone–burst frequency f_b and the corresponding pattern period T. The table shown below is our version of Table I in Pierce's paper.

Group # Pattern	Pattern frequency 1/T	Fundamental frequency (Hz)	Bursts per second	Cycles per burst n	Pattern period T(msec)
1a	3.90625	15.625	15.625	128	256
b	3.90625	3.90625	15.625	128	256
c	3.90625	3.90625	3.90625	128	256
2a	7.8125	31.25	31.25	64	128
b	7.8125	7.8125	31.25	64	128
c	7.8125	7.8125	7.8125	64	128
3a	15.625	62.5	62.5	32	64
b	15.625	15.625	62.5	32	64
c	15.625	15.625	15.625	32	64
4a	31.25	125	125	16	32
b	31.25	31.25	125	16	32
c	31.25	31.25	31.25	16	32
5a	62.5	250	250	8	16
b	62.5	62.5	250	8	16
c	62.5	62.5	62.5	8	16
6a	125	500	500	4	8
b	125	125	500	4	8
c	125	125	125	4	8
7a	250	1000	1000	2	4
b	250	250	1000	2	4
c	250	250	250	2	4
8a	500	2000	2000	1	2
b	500	500	2000	1	2
c	500	500	500	1	2

A set of eight test signals was generated according to this scheme. Each test signal consists of a sequence of the *a*, *b* and *c* patterns, each pattern lasting 1.024 seconds. This time interval was chosen to get an exact integer number of bursts, ranging from 4 for Case 1c to 2000 for Cases 8a and 8b.

The following sounds are presented in this video clip:

1) Pattern 1a
2) Pattern 8a (one cycle per burst)
3) Pattern 8b (one cycle per burst, every fourth pulse inverted)
4) Pattern 8c (one cycle per burst, 1/4 spacing of 8a)
5) Group Pattern 8 (1 cycle/burst)
6) Group Pattern 7 (2 cycles/burst)
7) Group Pattern 6 (4 cycles/burst)
8) Group Pattern 5 (8 cycles/burst)
9) Group Pattern 4 (16 cycles/burst)
10) Group Pattern 3 (32 cycles/burst)
11) Group Pattern 2 (64 cycles/burst)
12) Group Pattern 1 (128 cycles/burst)

Our conclusions (both auditory and visual) were very similar to those that Pierce reports, though the ranges were a bit different. We found very clear matching on pulse rate (*a* and *b*) for pattern frequencies of 62.5 Hz or less, and very clear matching on fundamental frequency (*b* and *c*) for pattern frequencies of 250 Hz or

more. The 125–Hz case was ambiguous, but could certainly have been called a pulse–rate match. The following table summarizes our admittedly subjective impressions.

Cycles/ burst n	Pattern freq 1/T	Matching patterns	Comments
128	3.90625	a b	a & b sound identical,“pulsy”; c much slower, same “tone” as a, b
64	7.8125	a b	a & b almost identical, “chopped”; c “pulsy,” same “tone” as a, b
32	15.625	a b	a & b very close, diff tone color; c “pulsy,” “tone” resembles a, b
16	31.25	a b	a & b close, harsh sounding; c “pulsy,” “tone” uncertain
8	62.5	a b	a & b related, b more inharmonic; c much lower in pitch, buzzy
4	125.	a~b	a & b related, a a bit higher pitch; c lower in pitch, buzzy
2	250.	b c	b & c very close, c buzzier; a two octaves higher in pitch
1	500.	b c	b & c identical

These tone–burst signals were processed by the cochlear model, and their correlograms were compared. While interpretation of the images is at least as subjective as interpretation of the sounds, there seemed to be a remarkable correspondence between the appearance of the correlograms and the results shown above.

For high pattern frequencies, the *b* and *c* correlograms were very similar, their common fundamental frequencies being clearly evident; they differed primarily in the *c* pattern having a cleaner high–frequency spectrum. This is roughly in accord with the fact that *b* and *c* have the same power spectrum. At the “critical” 125–Hz pattern frequency, the correlograms for *b* suddenly resembles *a* much more than *c*, a resemblance which becomes complete at the lower pattern frequencies. In none of these cases does one see much energy at the low fundamental frequencies, and the higher–frequency time structure of *b* dominates the match.

13. ASA Demonstrations (Starts at 25:10, length is 48:47)

The following examples are used with permission from the Acoustical Society of America. The sounds were published on ASA’s *Auditory Demonstrations* CD [Houtsma87]. A description of each of the following examples from the descriptive book that accompanies the CD can be found in an Appendix to this report. Note, the audio quality on a video tape is noticeably poorer than on the original compact disk.

The compact disk *Auditory Demonstrations* may be obtained from:

The Acoustical Society of America
500 Sunnyside Boulevard,
Woodbury, New York 10797.

Production of the disc was a joint effort of A.J.M. Houtsma and W.M. Wagenaars of the Institute of Perception Research in the Netherlands and T.D. Rossing of Northern Illinois University, with narration by I.J. Hirsh of the Central Institute for the Deaf in St. Louis.

13.1. Cancelled Harmonics (Starts at 25:10, length is 1:38)

13.9. Asymmetry of Masking by Pulsed Tones (Starts at 26:53, length is 1:37)

13.10. Backward and Forward Masking (Starts at 28:35, length is 1:25)

13.11. Pulsation Threshold (Starts at 33:07, length is 0:46)

13.12. Dependence of Pitch on Intensity (Starts at 34:13, length is 0:53)

13.13. Pitch Salience and Tone Duration (Starts at 35:13, length is 0:54)

13.14. Influence of Masking Noise on Pitch (Starts at 36:19, length is 0:32)

13.15. Octave Matching (Starts at 37:01, length is 1:51)

13.16. Stretched and Compressed Scales (Starts at 39:03, length is 1:03)

13.17. Frequency Difference Limen or JND (Starts at 40:16, length is 2:22)

13.18. Logarithmic and Linear Frequency Scales (Starts at 42:48, length is 0:25)

13.19. Pitch Streaming (Starts at 44:43, length is 1:26)

13.20. Virtual Pitch (Starts at 46:20, length is 0:43)

13.21. Shift of Virtual Pitch (Starts at 47:18, length is 1:15)

13.22. Masking Spectral and Virtual Pitch (Starts at 48:42, length is 1:35)

13.23. Virtual Pitch with Random Harmonic (Starts at 50:29, length is 1:09)

13.24. Strike Note of a Chime (Starts at 51:48, length is 1:28)

13.25. Analytic vs. Synthetic Pitch (Starts at 53:26, length is 0:32)

13.26. Scales with Repetition Pitch (Starts at 54:09, length is 1:31)

13.27. Circularity in Pitch Judgment (Starts at 55:52, length is 1:26)

13.28. Effect of Spectrum on Timbre (Starts at 57:30, length is 1:22)

13.29. Effect of Tone Envelope on Timbre (Starts at 59:05, length is 2:24)

13.30. Change in Timbre with Transposition (Starts at 1:01:39, length is 0:52)

13.31. Tones and Tuning with Stretched Partials (Starts at 1:02:41, length is 1:03)

13.32. Primary and Secondary Beats (Starts at 1:05:45, length is 1:38)

13.33. Distortion (Starts at 1:07:44, length is 2:24)

13.34. Aural Combination Tones (Starts at 1:10:18, length is 1:40)

13.35. Effect of Echoes (Starts at 1:12:06, length is 1:51)

14. Patterson Correlogram (Starts at 1:14:07, length is 0:54)

This clip compares a conventional correlogram with our implementation of Roy Patterson's Triggered Temporal Integration mechanism [Patterson91]. The Triggered Temporal Integration approach evolved from Patterson's Pulse Ribbon model [Patterson87] by adding a trigger mechanism that tries to find the pitch period of the signal. The goal of this mechanism is to create a stable image that shows the periodicities in the firing patterns at each position along the cochlea. In both cases, our second-order-transmission-line model of the cochlea [Slaney88] was used to model the transduction from acoustic pressure to nerve cell firings.

Unlike conventional correlograms, which summarize the temporal firing patterns of neurons by computing an autocorrelation, the Triggered Temporal Integration approach looks for landmarks in the firing pattern of each channel and synchronizes the display to these landmarks. The landmarks used are the peaks in the firing patterns and they fix the position of a period–synchronous integration. The firing probabilities in each channel are independently averaged on a period by period basis to form a stabilized image of the firing pattern.

Note, so that the correlograms are oriented the same way, we have flipped the horizontal axis of the Patterson correlogram. In Patterson's original implementation the trigger event is aligned along the right edge of the picture and the correlogram shows the time before the trigger. In our implementation the trigger event is along the left edge of the correlogram and time goes from right to left (or delay goes from left to right).

Our conclusion, based on this video clip and others, is that the triggered temporal integration scheme and conventional correlograms produce similar results. One difference between the two implementations is that an autocorrelation is symmetric so both negative and positive time lags are identical. You can see this difference when you look near the pitch period of steady state vowels.

Audio Commentary on Tape:

> "Hey Roy, Check it out! The top correlogram was computed using an implementation of your triggered temporal integration ideas. The bottom correlogram was computed using conventional autocorrelation. Let's see what vowels look like. First let's change the pitch:
>
> "Ahhh Ahhh Ahhh (with rising pitch from Clip 7)
>
> "Now, let's change the formants
>
> "Ahhh, Eeee, Iyyy, Ohhh, Oooh (from Clip 7)
>
> "Finally, let's see what happens with the famous Steve McAdams/Roger Reynolds oboe:
>
> (sound of oboe from Clip 4)
>
> "Looks, pretty good, doesn't it?"

15. Hydrodynamics Correlogram (Starts at 1:15:07, length is 0:54)

This clip compares a correlogram computed using the conventional transmission-line model used in the rest of the correlograms on this tape [Slaney88] versus a preliminary implementation of a new cochlear model based on the hydrodynamics of the cochlea [Lyon89]. The hydrodynamics model uses variable-Q filters to model the non-linear gains present in the cochlea. This clip shows the performance of the model over a 40dB range of input levels.

Audio Commentary on Tape:

"This is a test of the correlogram, comparing our new hydrodynamic model, on the right, with our previous cochlear model, on the left. Both models are being exercised at three loudness levels, differing by 20 dB increments, so that their gain-control responses may be compared. Let's see what vowels look like. First let's change the pitch:

"Ahhh Ahhh Ahhh (with rising pitch from Clip 7)

"Now, let's change the formants

"Ahhh, Eeee, Iyyy, Ohhh, Oooh (from Clip 7)

"Finally, let's see what happens with the famous Steve McAdams/Roger Reynolds oboe:

"(sound of oboe from Clip 4)"

References

Bregman90 — Albert S. Bregman, *Auditory Scene Analysis*, Bradford Books, MIT Press, Cambridge, MA, 1990.

Duda90 — Richard Duda, Richard Lyon and Malcolm Slaney, "Correlograms and the separation of sound," 24th Annual Asilomar Conference on Signals, Systems and Computers, Asilomar, CA, 1990.

Flanagan60 — James L. Flanagan and Newman Guttman, "Pitch of Periodic Pulses without Fundamental Component," *Journal of the Acoustical Society of America*, Vol. 32, pp. 1319-1328, 1960.

Horn86 — Berthold Klaus and Paul Horn, *Robot Vision*, The MIT Press, Cambridge, MA, 1986.

Houtsma87 — A. J. Houtsma, T. D. Rossing, and W. M. Wagenaars, *Auditory Demonstrations* (Compact Disc), Acoustical Society of America, (500 Sunnyside Boulevard, Woodbury, NY, 10797), 1987.

Licklider51 — J. C. R. Licklider, "A Duplex Theory of Pitch Perception," Experentia, Vol. 7, pp. 128-133, 1951. Also reprinted in *Psychological Acoustics*, E. D. Schubert (ed.), Dowden, Hutchinson and Ross, Inc., Stroudsburg, PA, 1979.

Lyon89 — Richard F. Lyon and Carver Mead, "Cochlear Hydrodynamics Demystified," Caltech Computer Science Technical Report Caltech–CS–TR–88–4, 1989.

McAdams84 — Stephen McAdams, "Spectral Fusion, Spectral Parsing and the Formation of Auditory Images," Technical Report STAN–M–22, Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, Stanford CA, May, 1984.

Meddis91 — Ray Meddis and Michael Hewitt, "Virtual Pitch and Phase Sensitivity of a Computer Model of the Auditory Periphery: I Pitch Indentification," submitted to *J. Acou. Soc. Am.*, 1991.

Mont–Reynaud89 — Bernard M. Mont–Reynaud and David K. Mellinger, "Source Separation by Frequency Co–Modulation," *Proc. First Int. Conf. Music Perception and Cognition*, pp. 99-102, Kyoto, Japan, October 1989.

Patterson87 — Roy D. Patterson, "A pulse ribbon model of monoural phase perception," *Journal of the Acoustical Society of America*, Vol. 82, pp. 1560-1586, 1987.

Patterson91 — Roy D. Patterson and John Holdsworth, "A functional model of neural activity patterns and auditory images," to appear in *Advances in Speech, Hearing and Language Processing, Volume 3*, edited by W. A. Ainsworth, JAI Press, London.

Pierce89 — John R. Pierce, "Toneburts: Explorations of Rate and Pitch Matching," internal note TONEB.NO8, Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, 1989.

Rabiner68 — Lawrence R. Rabiner, "Digital Formant Synthesizer for Speech–Synthesis Studies," *J. Acoust. Soc. Amer.*, Vol. 43, pp. 822-828, 1968.

Reynolds83 — Roger Reynolds, *Archipeligo*, for orchestra and computer–generated tape, C. F. Peters: N.Y., 1983

Schouten62 — J. F. Schouten, R. J. Ritsma, and B. Lopes Cardozo, "Pitch of the Residue," *Journal of the Acoustical Society of America*, Vol. 334, pp. 1418-14265, 1962.

Shephard64 — Roger N. Shepard, "Circularity in judgments of relative pitch," *J. Acoust. Soc. Am.*, 36, 2346-2353, 1964.

Slaney88 — Malcolm Slaney, "Lyon's Cochlear Model," Apple Computer Technical Report #13, Apple Computer, Inc., 1988 (this report is available from the Apple Corporate Library).

Slaney90 Malcolm Slaney and Richard F. Lyon, “A Perceptual Pitch Detector,” *Proceedings of the IEEE International Conference and Acoustics, Speech and Signal Processing* (ICASSP), Albuquerque, NM, March 1990.

Small70 Arnold W. Small, “Periodicity Pitch,” in *Foundations of Modern Auditory Theory*, Jerry V. Tobias (ed), Academic Press, New York, 1970.

Wolfram88 Stephen Wolfram, *Mathematica: A System for Doing Mathematics by Computer*, Addison-Wesley Publishing, Redwood City, CA, 1988.

Zue85 Victor Zue, “Use of Speech Knowledge in Automatic Speech Recognition,” *Proceedings of the IEEE*, Vol. 73, pp. 1602-1615, November 1985.

Appendix

ASA Auditory Demonstrations Descriptions

We have included correlograms of most of the examples on the ASA *Auditory Demonstrations* CD in this report. For the convenience of our viewers we have included the descriptions of these clips as part of this report. For the rest of the clips and the best quality audio we encourage you to purchase the CD directly from the Acoustical Society. The CD should be part of everybody's library.

1. Cancelled Harmonics (Starts at 25:10, length is 1:38)

This demonstration illustrates Fourier analysis of a complex tone consisting of 20 harmonics of a 200–Hz fundamental. The demonstration also illustrates how our auditory system, like our other senses, has the ability to listen to complex sounds in different modes. When we listen analytically, we hear the different components separately; when we listen holistically, we focus on the whole sound and pay little or no attention to the components.

When the relative amplitudes of all 20 harmonics remain steady (even if the total intensity changes), we tend to hear them holistically. However, when one of the harmonics is turned off and on, it stands out clearly. The same is true if one of the harmonics is given a "vibrato" (i.e., its frequency, its amplitude, or its phase is modulated at a slow rate).

Commentary

> "A complex tone is presented, followed by several cancellations and restorations of a particular harmonic. This is done for harmonics 1 through 10."

References

- R.Plomp (1964), "The ear as a frequency analyzer," J. Acoust. Soc. Am. 36, 1628-367.
- H. Duifhuis (1970), "Audibility of high harmonics in a periodic pulse," J. Acoust. Soc. Am. 48, 888-93.

9. Asymmetry of Masking by Pulsed Tones (Starts at 26:53, length is 1:37)

A pure tone masks tones of higher frequency more effectively than tones of lower frequency. This may be explained by reference to the simplified response of the basilar membrane for two pure tones A and B shown in the figure below. In (a), the excitations barely overlap; little masking occurs. In (b) there is appreciable overlap; tone B masks tone A more than A masks B. In (c) the more intense tone B almost completely masks the higher–frequency tone A. In (d) the more intense tone A does not mask the lower–frequency tone B.

This demonstration uses tones of 1200 and 2000 Hz, presented as 200–ms tone bursts separated by 100 ms (see figure above). The test tone, which appears every other pulse, decreases in 10 steps of 5 dB each, except the first step which is 15 dB.

Commentary

> "A masking tone alternates with the combination of masking tone plus a stepwise–decreasing test tone. First the masker is 1200 Hz and the test tone is 2000 Hz, then the masker is 2000 Hz and the test tone is 1200 Hz. Count how many steps of the test tone can be heard in each case."

References

- G.von Bekesy (1970), "Traveling waves as frequency analyzers in the cochlea," Nature 225, 1207-09.
- J.P.Egan and H.W.Hake (1950), "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. 22, 622-30.

- R.Patterson and D.Green (1978), "Auditory masking," in *Handbook of Perception*, Vol. 4: *Hearing*, ed. E.Carterette and M.Friedman (Academic Press, New York) pp. 337-62.
- T.D.Rossing (1982), *The Science of Sound* (Addison-Wesley, Reading, MA). Chap.
- J.J.Zwislocki (1978), "Masking: Experimental and theoretical aspects of simultaneous, forward, backward, and central masking," in *Handbook of Perception*, Vol. 4: *Hearing*, ed. E.Carterette and M.Friedman (Academic Press, New York) pp. 283-336.

10. Backward and Forward Masking (Starts at 28:35, length is 1:25)

Masking can occur even when the tone and the masker are not simultaneous. Forward masking refers to the masking of a tone by a sound that ends a short time (up to about 20 or 30 ms) before the tone begins. Forward masking suggests that recently stimulated sensors are not as sensitive as fully–rested sensor. Backward masking refers to the masking of a tone by a sound that begins a few milliseconds after the tone has ended. A tone can be masked by noise that begins up to 10 ms later, although the amount of masking decreases as the time interval increases (Elliot, 1962). Backward masking apparently occurs at higher centers of processing in the nervous system where the neural correlates of the later–occurring stimulus of greater intensity overtake and interfere with those of the weaker stimulus.

First the signal (10–ms bursts of a 2000–Hz sinusoid) is presented in 10 decreasing steps of –4 dB without a masker. Next, the 2000–Hz signal is followed after a time gap, t, by 250–ms bursts of noise (1900–2100 Hz). The time gap, t, is successively 100 ms, 20 ms, and 0. The sequence is repeated.

Finally, the masker is presented before the tone, again with t = 100 ms, 20 ms, and 0.

Commentary

> "First you will hear a brief sinusoidal tone, decreasing in 10 steps of 4 decibels each.
>
> "Now the same signal is followed by a noise burst with a brief time gap in between. It is heard alternating with the noise burst alone. For three decreasing time–gap values, you will hear two staircases. Count the number of steps for which you can hear the brief signal preceding the noise.
>
> "Now the noise burst precedes the signal. Again two staircases are heard for each of the same three time–gap values. Count the number of steps that you can hear the signal following the noise."

References

- H.Duifhuis (1973), "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. 54, 1471-88.
- L.L.Elliot (1962), "Backward and forward masking of probe tones of different frequencies," J. Acoust. Soc. Am. 34, 1116-17.
- J.H.Patterson (1971), "Additivity of forward and backward masking as a function of signal frequency," J. Acoust. Soc. Am. 50, 1123-25.

11. Pulsation Threshold (Starts at 33:07, length is 0:46)

Perception (e.g.,visual, auditory) is an interpretive process. If our view of one object is obscured by another, for example, our perception may be that of two intact objects even though this information is not present in the visual image. In general, our interpretive processes provide us with an accurate picture of the world; occasionally, they can be fooled (e.g., visual or auditory illusions).

Such interpretive processes can be demonstrated by alternating a sinusoidal signal with bursts of noise. Whether the signal is perceived as pulsating or continuous depends upon the relative intensities of the signal and noise.

In this demonstration, 125–ms bursts of a 2000–Hz tone alternate with 125–ms bursts of noise (1875–2125 Hz), as shown below. The noise level remains constant, while the tone level decreases in 15 steps of –l dB after each 4 tones.

The pulsation threshold is given by the level at which the 2000–Hz to sound continuous.

Commentary

> "You will hear a 2000–Hz tone alternating with a band of noise centered around 2000 Hz. The tone intensity decreases one decibel after every four tone presentations. Notice when the tone begins to appear continuous."

References

- A.S.Bregman (1978), "Auditory streaming: competition among alternative organizations," Percept. Psychophys. 23, 391-98.
- T.Houtgast (1972), "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am. 51, 1885-94.

12. Dependence of Pitch on Intensity (Starts at 34:13, length is 0:53)

Early experimenters reported substantial pitch dependence on intensity. Stevens (1935), for example, reported apparent frequency changes as large as 12% as the sound level of sinusoidal tones increased from 40 to 90 dB. It now appears that the effect is small and varies considerably from subject to subject. Whereas Terhardt (1974) found pitch changes for some individuals as large as those reported by Stevens, averaging over a group of observers made them insignificant.

Using tones of long duration, Stevens (1935) found that tones below 1000 Hz decrease in apparent pitch with increasing intensity, whereas tones above 2000 Hz increase their pitch with increasing intensity. Using 40–ms bursts, however, Rossing and Houtsma (1986) found a monotonic decrease in pitch with intensity over the frequency range 200–3200 Hz, as did Doughty and Garner (1948) using 12–ms bursts.

In the demonstration, we use 500–ms tone bursts having frequencies of 200, 500,1000, 3000, and 4000 Hz. Six pairs of tones are presented at each frequency, with the second tone having a level that is 30 dB higher than the first one (which is 5 dB above the 200–Hz calibration tone). For most pairs, a slight pitch change will be audible.

Commentary

> "First, a 200–Hz calibration tone. Adjust the level so that it is just audible.
>
> "Now, 6 tone pairs are presented at various frequencies. Compare the pitches for each tone pair."

References

- J.M.Doughty and W.M.Garner (1948), "Pitch characteristics of short tones II: Pitch as a function of duration," J. Exp. Psychol. 38, 478-94.
- T.D.Rossing and A.J.M.Houtsma (1986), "Effects of signal envelope on the pitch of short sinusoidal tones," J. Acoust. Soc. Am. 79, 1926-33.
- S.S.Stevens (1935), "The relation of pitch to intensity," J. Acoust. Soc. Am. 6, 150-54 .
- E.Terhardt (1974), "Pitch of pure tones: its relation to intensity," in *Facts and Models in Hearing*, ed. E.Zwicker and E.Terhardt (Springer Verlag, New York) pp. 353-60.
- J.Verschuure and A.A.van Meeteren (1975), "The effect of intensity on pitch," Acustica 32, 33-44.

13. Pitch Salience and Tone Duration (Starts at 35:13, length is 0:54)

How long must a tone be heard in order to have an identifiable pitch? Early experiments by Savart (1830) indicated that a sense of pitch develops after only two cycles. Very brief tones are described as "clicks," but as the tones lengthen, the clicks take on a sense of pitch which increases upon further lengthening.

It has been suggested that the dependence of pitch salience on duration follows a sort of "acoustic uncertainty principles,"

$$\Delta f \ \Delta t = K,$$

where Δf is the uncertainty in frequency and Δt is the duration of a tone burst. K, which can be as short as 0.1 (Majernik and Kaluzny, 1979), appears to depend upon intensity and amplitude envelope (Ronken,

1971). The actual pitch appears to have little or no dependence upon duration (Doughty and Garner, 1948; Rossing and Houtsma, 1986).

In this demonstration, we present tones of 300, 1000, and 3000 Hz in bursts of 1, 2, 4, 8, 16, 32, 64, and 128 periods. How many periods are necessary to establish a sense of pitch?

Commentary

> "In this demonstration, three tones of increasing durations are presented. Notice the change from a click to a tone. Sequences are presented twice."

References

- J.M.Doughty and W.M.Garner (1948), "Pitch characteristics of short tones II: pitch as a function of duration," J. Exp. Psych. 38, 478-94.
- V.Majernik and J.Kaluzny (1979), "On the auditory uncertainty relations," Acustica 43, 132-46.
- D.A.Ronken (1971), "Some effects of bandwidth–duration constraints on frequency discrimination," J. Acoust. Soc. Am. 49 1232-42.
- T.D.Rossing and A.J.M.Houtsma (1986), "Effects of signal envelope on the pitch of short sinusoidal tones," J. Acoust. Soc. Am. 79, 1926-33.
- F.Savart (1830), "Uber die Ursachen der Tonhohe," Ann. Phys. Chem. 51, 555.

14. Influence of Masking Noise on Pitch (Starts at 36:19, length is 0:32)

The pitch of a tone is influenced by the presence of masking noise or another tone near to it in frequency. If the interfering tone has a lower frequency, an upward shift in the test tone is always observed. If the interfering tone has a higher frequency, a downward shift is observed, at least at low frequency (< 300 Hz). Similarly, a band of interfering noise produces an upward shift in a test tone if the frequency of the noise is lower (Terhardt and Fastl, 1971).

In this demonstration, a 1000–Hz tone, 500 ms in duration and partially masked by noise low–pass filtered at 900 Hz, alternates with an identical tone, presented without masking noise. The tone partially masked by noise of lower frequency appears slightly higher in pitch (do you agree?). When the noise is turned off, it is clear that the two tones were identical.

Commentary

> "A partially masked 1000–Hz tone alternates with an unmasked 1000–Hz comparison tone. Compare the pitches of the two tones."

References

- B.Scharf and A.J.M.Houtsma (1986), "Audition II: Loudness, pitch, localization, aural distortion, pathology," in *Handbook of Perception and Human Performance*, Vol. 1, ed. K.R.Boff, L.Kaufman, and J.P.Thomas (J. Wiley, New York).
- E.Terhardt and H.Fastl (1971), "Zum Einfluss von Stortonen und Storgerauschen auf die Tonhohe von Sinustonen," Acustica 25, 53-61.

15. Octave Matching (Starts at 37:01, length is 1:51)

Experiments on octave matching usually indicate a preference for ratios that are greater than 2.0. This preference for stretched octaves is not well understood. It is only partly related to our experience with hearing stretch–tuned pianos. More likely, it is related to the phenomenon we encountered in Demonstration 14, although in this demonstration the tones are presented alternately rather than simultaneously.

In this demonstration, a 500–Hz tone of one second duration alternates with another tone that varies from 985 to 1035 Hz in steps of 5 Hz. Which one sounds like a correct octave? Most listeners will probably select a tone somewhere around 1010 Hz.

Commentary

> "A 500–Hz tone alternates with a stepwise increasing comparison tone near 1000 Hz. Which step seems to represent a "correct" octave? The demonstration is presented twice."

References

- D.Allen (1967), “Octave discriminibility of musical and non–musical subjects,” Psychonomic Sci. 7, 421-22.
- E.M.Burns and W.D.Ward (1982), “Intervals, scales, and tuning,” in *The Psychology of Music*, ed. D.Deutsch (Academic Press, New York) pp. 241-69.
- J.E.F.Sundberg and J.Lindqvist (1973), “Musical octaves and pitch,” J. Acoust. . Soc. Am. 54, 922-29.
- W.D.Ward (1954), “Subjective musical pitch,” J. Acoust. Soc. Am. 26, 369-80.

16. Stretched and Compressed Scales (Starts at 39:03, length is 1:03)

This demonstration, for which we are indebted to E.Terhardt, illustrates that to many listeners an over–stretched intonation, such as case (b) is acceptable, whereas a compressed intonation (a) is not. Terhardt has found that about 40% of a large audience will judge intonation (b) superior to the other two. The program is as follows:

a) intonation compressed by a semitone (bass in C, melody in B);

b) intonation stretched by a semitone (bass in C, melody in C#);

c) intonation “mathematically correct” (bass and melody in C).

Commentary

> “You will hear a melody played in a high register with an accompaniment in a low register. Which of the three presentations sounds best in tune?”

In case you wish to sing along, here are some words to go with the melody:

> In Munchen steht ein Hofbrauhaus, eins, zwei gsuffa
> Da lauft so manches Wasserl aus, eins, zwei, gsuffa
> Da hat so mancher brave Mann, eins, zwei, gsuffa
> Gezeigt was er vertragen kann,
> Schon fruh am Morgen fangt er an
> Und spat am Abend hort er auf
> So schon ist’s im Hofbrauhaus!
> (authenticity not guaranteed)

References

- E.Terhardt and M.Zick (1975), “Evaluation of the tempered tone scale in normal, stretched, and contracted intonation,” Acustica 32, 268-74.
- E.M.Burns and W.D.Ward (1982), “Intervals, scales, and tuning,” in *The Psychology of Music*, ed. D.Deutsch (Academic Press, New York) pp. 241-69.

17. Frequency Difference Limen or JND (Starts at 40:416, length is 2:22)

The ability to distinguish between two nearly equal stimuli is often characterized by a difference limen (DL) or just noticeable difference (jnd). Two stimuli cannot be consistently distinguished from one another if they differ by less than a jnd.

The jnd for pitch has been found to depend on the frequency, the sound level, the duration of the tone, and the suddenness of the frequency change. Typically, it is found to be about 1/30 of the critical bandwidth at the same frequency.

In this demonstration, 10 groups of 4 tone pairs are presented. For each pair, the second tone may be higher (A) or lower (B) than the first tone. Pairs are presented in random order within each group, and the frequency difference decreases by 1 Hz in each successive group. The tones, 500 ms long, are separated by 250 ms. Following is the order of pairs within each group, where A represents (f,f+ Δf), B represents (f+ Δf,f), and f equals 1000 Hz:

Group	Df (Hz)	Key
1	10	A,B,A,A
2	9	A,B,B,B
3	8	B,A,A,B
4	7	B,A,A,B
5	6	A,B,A,B
6	5	A,B,A,A
7	4	B,B,A,A
8	3	A,B,A,B
9	2	B,B,B,A
10	1	B,A,A,B

Commentary

> "You will hear ten groups of four tone pairs. In each group there is a small frequency difference between the tones of a pair, which decreases in each successive group."

References

- B.C.J.Moore (1974), "Relation between the critical bandwidth and the frequency difference limen," J. Acoust. Soc. Am. 55, 359.
- C.C.Wier, W.Jesteadt, and D.M.Green (1977), "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. 61, 178-84.
- E.Zwicker (1970), "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, ed. R.Plomp and G.F.Smoorenburg (Sijthoff, Leiden).

18. Logarithmic and Linear Frequency Scales (Starts at 42:48, length is 0:25)

A musical scale is a succession of notes arranged in ascending or descending order. Most musical composition is based on scales, the most common ones being those with five notes (pentatonic), twelve notes (chromatic), or seven notes (major and minor diatonic, Dorian and Lydian modes, etc.). Western music divides the octave into 12 steps called semitones. All the semitones in an octave constitute a chromatic scale or 12–tone scale. However, most music makes use of a scale of seven selected notes, designated as either a major scale or a minor scale and carrying the note name of the lowest note. For example, the C–major scale is played on the piano by beginning with any C and playing white keys until another C is reached.

Other musical cultures use different scales. The pentatonic or five–tone scale, for example, is basic to Chinese music but also appears in Celtic and Native American music. A few cultures, such as the Nasca Indians of Peru, have based their music on linear scales (Haeberli, 1979), but these are rare. Most music is based on logarithmic (steps of equal frequency ratio $\Delta f/f$) rather than linear (steps of equal frequency Δf) scales .

In this demonstration we compare both 7–step diatonic and 12–step chromatic scales with linear and logarithmic steps.

Commentary

> "Eight–note diatonic scales of one octave are presented. Alternate scales have linear and logarithmic steps. The demonstration is repeated once.
>
> "Next, 13–note chromatic scales are presented, again alternating between scales with linear and logarithmic steps."

References

- A.H.Benade (1976), *Fundamentals of Musical Acoustics* (Oxford Univ., New York). Chap. 15.
- E.M.Burns and W.D.Ward (1982), "Intervals, scales, and tuning," in *The Psychology of Music*, ed. D.Deutsch (Academic Press, New York). pp. 241-69.
- J.Haeberli (1979), "Twelve Nasca panpipes: A study," Ethomusicology 23, 57-74.

- D.E.Hall (1980), *Musical Acoustics* (Wadsworth, Belmont, CA) pp. 444-51.
- T.D.Rossing (1982), *The Science of Sound* (Addison-Wesley, Reading, MA) Chap. 9

19. Pitch Streaming (Starts at 44:43, length is 1:26)

It is clear in listening to melodies that sequences of tones can form coherent patterns. This is called temporal coherence. When tones do not form patterns, but seem isolated, that is called fission.

Temporal coherence and fission are illustrated in a demonstration first presented by van Noorden (1975) and included in the "Harvard tapes" (1978). Van Noorden describes it as a "galloping rhythm."

We present tones A and B in the sequence ABA ABA. Tone A has a frequency of 2000 Hz, tone B varies from 1000 to 4000 Hz and back again to 1000 Hz. Near the crossover points, the tones appear to form a coherent pattern, characterized by a galloping rhythm, but at large intervals the tones seem isolated, illustrating fission.

Commentary

> "In this experiment a fixed tone A and a variable tone B alternate in a fast sequence ABA ABA. At some places you may hear a "galloping rhythm," while at other places the sequences of tone A and B seem isolated."

References

- A.S.Bregman (1978), "Auditory streaming: competition among alternative organizations," Percept. Psychophys. 23, 391-98.
- L.P.A.S.van Noorden (1975), *Temporal Coherence in the Perception of Tone Sequences*. Doctoral dissertation with phonograph record (Institute for Perception Research, Eindhoven, The Netherlands).
- Harvard University Laboratory of psychophysics (1978), "Auditory demonstration tapes," No. 18.

20. Virtual Pitch (Starts at 46:20, length is 0:43)

A complex tone consisting of 10 harmonics of 200 Hz having equal amplitude is presented, first with all harmonics, then without the fundamental, then without the two lowest harmonics, etc. Low–frequency noise (300–Hz lowpass, –10 dB) is included to mask a 200–Hz difference tone that might be generated due to distortion in playback equipment.

Commentary

> "You will hear a complex tone with 10 harmonics, first complete and then with the lower harmonics successively removed. Does the pitch of the complex change? The demonstration is repeated once."

References

- .J.M.Houtsma and J.L.Goldstein (1972), "The central origin of the pitch of complex tones: evidence from musical interval recognition," J. Acoust. Soc. Am. 51, 520-529.
- J.F.Schouten (1940), "The perception of subjective tones," Proc. Kon. Ned. Akad. Wetenschap 41, 1086-1093.
- A.Seebeck (1841), "Beobachtungen uber einige Bedingungen der Entstehung von Tonen," Ann. Phys. Chem. 53, 417-436.

21. Shift of Virtual Pitch (Starts at 47:18, length is 1:15)

A tone having strong partials with frequencies of 800, 1000, and 1200 Hz will have a virtual pitch corresponding to the 200 Hz missing fundamental, as in Demonstration 20. If each of these partials is shifted upward by 20 Hz, however, they are no longer exact harmonics of any fundamental frequency around 200 Hz. The auditory system will accept them as being "nearly harmonic" and identify a virtual pitch slightly above 200 Hz (approximately $1/3(820/4 + 1020/5 + 1220/6) = 204$ Hz in this case). The auditory system appears to search for a "nearly common factor" in the frequencies of the partials.

Note that if the virtual pitch were created by some kind of distortion, the resulting difference tone would remain at 200 Hz when the partials were shifted upward by the same amount.

In this demonstration, the three partials in a complex tone, 0.5 s in duration, are shifted upward in ten 20–Hz steps while maintaining a 200–Hz spacing between partials. You will almost certainly hear a virtual pitch that rises from 200 to about 1/3(1000/4 + 1200/5 + 1400/6) = 241 Hz. At the same time, you may have noticed a second rising virtual pitch that ends up at 1/3(1000/5 + 1200/6 + 1400/7) = 200 Hz and possibly even a third one, as shown in Fig. 2 in Schouten et al. (1962)

In the second part of the demonstration it is shown that virtual pitches of a complex tone having partials of 800, 1000, and 1200 Hz and one having partials of 850, 1050, and 1250 Hz can be matched to harmonic complex tones with fundamentals of 200 and 210 Hz respectively.

Commentary

> "You will hear a three–tone harmonic complex with its partials shifted upward in equal steps until the complex is harmonic again. The sequence is repeated once.
>
> "Now you hear a three–tone complex of 800, 1000 and 1200 Hz, followed by a complex of 850, 1050 and 1250 Hz. As you can hear, their virtual pitches are well matched by the regular harmonic tones with fundamentals of 200 and 210 Hz. The sequence is repeated once."

References

- J.F.Schouten, R.L.Ritz and B.L.Cardozo (1962), "Pitch of the residue," J. Acoust. Soc. Am. 34, 1418-1424.
- G.F.Smoorenburg (1970), "Pitch perception of two–frequency stimuli," J. Acoust. Soc. Am. 48, 926-942.

22. Masking Spectral and Virtual Pitch (Starts at 48:42, length is 1:35)

This demonstration uses masking noise of high and low frequency to mask out, alternately, a melody carried by single pure tones of low frequency and the same melody resulting from virtual pitch from groups of three tones of high frequency (4th, 5th and 6th harmonics). The inability of the low–frequency noise to mask the virtual pitch in the same range points out the inadequacy of the place theory of pitch.

Commentary

> "You will hear the familiar Westminster chime melody played with pairs of tones. The first tone of each pair is a sinusoid, the second a complex tone of the same pitch."
>
> "Now the pure tone notes are masked with low–pass noise. You will still hear the pitches of the complex tone."
>
> "Finally the complex tone is masked by high–pass noise. The pure–tone melody is still heard."

References

- J.C.R.Licklider (1955), "Influence of phase coherence upon the pitch of complex tones," J. Acoust. Soc. Am. 27, 996 (A).
- R.J.Ritsma and B.L.Cardozo (1963/64), "The perception of pitch," Philips Techn. Review 29, 37-43.

23. Virtual Pitch with Random Harmonic (Starts at 50:29, length is 1:09)

This demonstration compares the virtual pitch and the timbre of complex tones with low and high harmonics of a missing fundamental. Within three groups (low, medium, high harmonics) the harmonic numbers are randomly chosen. The Westminster chime melody is clearly recognizable in all three cases. This shows that the recognition of a virtual–pitch melody is not based on aural tracking of a particular harmonic in the successive complex–tone stimuli.

Commentary

> "The tune of the Westminster chime is played with tone complexes of three random successive harmonics. In the first presentation harmonic numbers are limited between 2 and 6.
>
> "Now harmonic numbers are between 5 and 9.
>
> "Finally harmonic numbers are between 8 and 12."

References

- A.J.M.Houtsma (1984), "Pitch salience of various complex sounds," Music Perc. 1, 296-307.

24. Strike Note of a Chime (Starts at 51:48, length is 1:28)

In a carillon bell or a tuned church bell, the pitch of the strike note nearly coincides with the second partial (called the Fundamental or Prime), and thus the two are difficult to separate. In most orchestral chimes, however, the strike note lies between the second and third partial (Rossing, 1982). Its pitch is usually identified as the missing fundamental of the 4th, 5th and 6th partials, which have frequencies nearly in the ratio 2:3:4. A few listeners identify the chime strike note as coinciding with the 4th partial (an octave higher). In which octave do you hear it?

Commentary

> "An orchestral chime is struck eight times, each time preceded by cue tones equal to the first eight partials of the chime.
>
> "The chime is now followed by a tone matching its nominal pitch or strike notes."

References

- T.D.Rossing (1982) *The Science of Sound*, (Addison-Wesley, Reading MA), pp. 241-242.

25. Analytic vs. Synthetic Pitch (Starts at 53:26, length is 0:32)

Our auditory system has the ability to listen to complex sounds in different modes. When we listen analytically, we hear the different frequency components separately; when we listen synthetically or holistically, we focus on the whole sound and pay little attention to its components.

In this demonstration, which was originally described by Smoorenburg (1970), a two–tone complex of 800 and 1000 Hz is followed by one of 750 and 1000 Hz. If you listen analytically, you will hear one partial go down in pitch; if you listen synthetically you will hear the pitch of the missing fundamental go up in pitch (a major third, from 200 to 250 Hz).

Commentary

> "A pair of complex tones is played four times against a background of noise. Do you hear the pitch go up or down?"

References

- H.L.F. von Helmholtz (1877), *On the Sensation of Tone*, 4th ed. Transl., A.J.Ellis, (Dover, New York, 1954).
- G.F.Smoorenburg (1970), "Pitch perception of two–frequency stimuli," J. Acoust. Soc. Am. 98, 924-942.
- E.Terhardt (1974), "Pitch, consonance and harmony," J. Acoust. Soc. Am. 55, 1061-1069.

26. Scales with Repetition Pitch (Starts at 54:09, length is 1:31)

One way to demonstrate repetition pitch in a convincing way is to play scales or melodies using pairs of pulses with appropriate time delays between members of a pair.

In the first demonstration, a 5–octave diatonic scale is presented using pairs of identical pulses. In each pair the second pulse repeats the first, with a certain time delay, which changes from 15 ms to 0.48 ms as we proceed up the scale.

Next a 4–octave diatonic scale is presented using trains of pulse pairs that occur at random times. Time intervals between leading pulses have a Poisson distribution. Delay times tau between pulses of each pair vary from 15 to 0.85 ms. Although the pulses now appear at random times, a pitch corresponding to l/t is heard, as in the first demonstration.

The final demonstration is a 4–octave diatonic scale played with bursts of white noise that are echoed 15 ms to 0.85 ms later.

Commentary

> "First you will hear a 5–octave diatonic scale played with pulse pairs.
>
> "Now you hear a 4–octave diatonic scale played with pulse pairs that are samples of a Poisson process.
>
> "Finally you hear a 4–octave diatonic scale played with bursts of echoed or 'comb filtered' white noise."

References

- F.A.Bilsen (1966), "Repetition pitch: monaural interaction of a sound with repetition of the same, but phase shifted sound", Acustica 17, 295-300.

27. Circularity in Pitch Judgment (Starts at 55:52, length is 1:26)

One of the most widely used auditory illusions is Shepard's (1964) demonstration of pitch circularity, which has come to be known as the "Shepard Scale" demonstration. The demonstration uses a cyclic set of complex tones, each composed of 10 partials separated by octave intervals. The tones are cosinusoidally filtered to produce the sound level distribution shown below, and the frequencies of the partials are shifted upward in steps corresponding to a musical semitone (≈6%). The result is an "ever–ascending" scale, which is a sort of auditory analog to the ever–ascending staircase visual illusion.

Several variations of the original demonstration have been described. J–C. Risset created a continuous version. Other variations are described by Burns (1981), Teranishi (1986), and Schroeder (1986).

Commentary

> "Two examples of scales that illustrate circularity in pitch judgment are presented. The first is a discrete scale of Roger N. Shepard, the second a continuous scale of Jean–Claude Risset."

References

- R.N.Shepard (1964), "Circularity in judgments of relative pitch," J. Acoust. Soc. Am. 36, 2346-2353.
- I.Pollack (1977), "Continuation of auditory frequency gradients across temporal breaks," Perception and Psychophys. 21, 563-568.
- E.M.Burns (1981) "Circularity in relative pitch judgments: the Shepard demonstration revisited, again," Perception and Psychophys. 30, 467-472.
- R.Teranishi (1986), "Endlessly rising or falling chordal tones which can be played on the piano; another variation of the Shepard demonstration," Paper K2-3, 12th Int. Congress on Acoustics, Toronto.
- M.R.Schroeder (1986),"Auditory paradox based on fractal waveform," J. Acoust. Soc. Am. 79, 186-189.

28. Effect of Spectrum on Timbre (Starts at 57:30, length is 1:22)

The sound of a Hemony carillon bell, having a strike–note pitch around 500 Hz (B4), is synthesized in eight steps by adding successive partials with their original frequency, phase and temporal envelope. The partials added in successive steps are:

1. Hum note (251 Hz)
2. Prime or Fundamental (501 Hz)
3. Minor Third and Fifth (603, 750 Hz)
4. Octave or Nominal (1005 Hz)
5. Duodecime or Twelfth (1506 Hz)
6. Upper Octave (2083 Hz)
7. Next two partials (2421, 2721 Hz)
8. Remainder of partials

The sound of a guitar tone with a fundamental frequency of 251 Hz is analyzed and resynthesized in a similar manner. The partials added in successive steps are:

1. Fundamental
2. 2nd harmonic
3. 3rd harmonic
4. 4th harmonic
5. 5th and 6th harmonics
6. 7th and 8th harmonics
7. 9th, 10th and 11th harmonics
8. Remainder of partials

Commentary

> "You will hear the sound of two instruments built up by adding partials one at a time."

References

- A. Lehr (1986), "Partial groups in the bell sound," J. Acoust. Soc. Am. 79, 2000-2011.
- J.C. Risset and M.V. Mathews (1969), "Analysis of musical instrument tones," Physics Today 22, 23-30.

29. Effect of Tone Envelope on Timbre (Starts at 59:05, length is 2:24)

The purpose of this demonstration (originally presented by J. Fassett) is to show that the temporal envelope of a tone, i.e. the time course of the tone's amplitude, has a significant influence on the perceived timbre of the tone. A typical tone envelope may include an attack, a steady–state, and a decay portion (e.g., wind instrument tones), or may merely have an attack immediately followed by a decay portion (e.g., plucked or struck string tones). By removing the attack segment of an instrument's sound, or by substituting the attack segment of another musical instrument, the perceived timbre of the tone may change so drastically that the instrument is no longer recognizable.

In this demonstration, a four–part chorale by J.S. Bach ("Als der gutige Gott") is played on a piano and recorded on tape. Next the chorale is played backward on the piano from end to beginning, and recorded again. Finally the tape recording of the backward chorale is played in reverse, yielding the original (forward) chorale, except that each note is reversed in time. The instrument does not sound like a piano any more, but rather resembles a kind of reed organ. The power spectrum of each note, measured over the note's duration, is not changed by temporal reversal of the tone.

Commentary

> "You will hear a recording of a Bach chorale played on a piano.
>
> "Now the same chorale will be played backwards.
>
> "Now the tape of the last recording is played backwards so that the chorale is heard forward again, but with an interesting difference."

References

- K.W.Berger (1963), "Some factors in the recognition of timbre," J. Acoust. Soc. Am. 36, 1888-1891.

- M.Clark, D.Luce, R.Abrams, H.Schlossberg and J.Rome (1964), “Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones,” J. Audio Eng. Soc. 12, 28-31.
- J.Fassett, “Strange to your Ears,” Columbia Record No. ML 4938.

30. Change in Timbre with Transposition (Starts at 1:01:39, length is 0:52)

High and low tones from a musical instrument normally do not have the same relative spectrum. A low tone on the piano typically contains little energy at the fundamental frequency and has most of its energy at higher harmonics. A high piano tone, however, typically has a strong fundamental and weaker overtones.

If a single tone from a musical instrument is spectrally analyzed and the resulting spectrum is used as a model for the other tones, one almost always obtains a series of tones that do not seem to come from that instrument. This is demonstrated with a recorded 3–octave diatonic scale played on a bassoon. A similar 3–octave “bassoon” scale is then synthesized by temporal stretching of the highest tone to obtain the proper pitch for each lower note on the scale. Segments of the steady–state portions are removed to retain the original note lengths. This way the spectra of all tones on the scale are identical except for a frequency scale factor. The listener will notice that, except for the highest note, the second scale does not sound as played on a bassoon.

Commentary

> “A 3–octave scale on a bassoon is presented, followed by a 3–octave scale of notes that are simple transpositions of the instrument’s highest tone. This is how the bassoon would sound if all its tones had the same relative spectrum.”

References

- J.Backus (1977), *The Acoustical Foundations of Music*, 2nd edition, (Norton &Co., New York).
- P.R.Lehman (1962), *The Harmonic Structure of the Tone of the Bassoon*, (Ph.D. Dissertation, Univ. of Mich.).

31. Tones and Tuning with Stretched Partials (Starts at 1:02:41, length is 1:03)

Most tonal musical instruments used in Western culture have spectra that are exactly or nearly harmonic. Tone scales used in Western music are also based on intervals or frequency ratios of simple integers, such as the octave (2:1), the fifth (3:2), and the major third (5:4). When several instruments play a consonant chord such as a major triad (do–mi–so), harmonics will match so that no beats will occur. If one changes the melodic scale to, for instance, a scale of equal temperament with 13 tones in an octave, a normally consonant triad may cause an unpleasant sensation of beats because some harmonics will have nearly but not exactly the same frequency. Similarly, instruments with a nonharmonic overtone structure, such as conventional carillon bells, can create unpleasant beat sensations even if their pitches are tuned to a natural scale. Beats will not occur, however, if the melodic scale of an instrument’s tones matches the overtone structure of those tones.

First, a four–part chorale (“Als der gutige Gott”) by J.S. Bach is played on a synthesized piano–like instrument whose tones have s exactly harmonic partials with amplitudes inversely proportional to harmonic number, and with exponential time decay. The melodic scale used is equally tempered, with semitone frequency ratios of the 12th root of 2.

In the second example, the same piece is played with equally stretched harmonic as well as melodic frequency ratios. The harmonics of each tone have been uniformly stretched on a log–frequency scale such that the second harmonic is 2.1 times the fundamental frequency, the 4th harmonic 4.41 times the fundamental, etc. The melodic scale is similarly tuned in such a way that each “semitone” step represents a frequency ratio of the 12th root of 2.1. The music is, in a sense, not dissonant because no beats occur. Nevertheless the harmonies may sound less consonant than they did in the first demonstration. This suggests that the presence or absence of beats is not the only criterion for consonance. The listener may also find it difficult to tell how many voices or parts the chorale has, since notes seem to have lost their gestalt due to the inharmonicity of their partials.

In the third example, the tones are made exactly harmonic again, but the melodic scale remains stretched to an "octave" ratio of 2.1. Disturbing beats are heard, but the four voices have regained their gestalt. The piece sounds as if it is played on an out–of–tune instrument.

In the final example, the harmonics of all tones are stretched, as was done in example 2, but the melodic scale is one of equal temperament based on an octave ratio of 2.0. Again there are annoying beats. This time, however, it is again very difficult to hear how many voices the chorale has.

Commentary

> "You will hear a 4–part Bach chorale played with tones having 9 harmonic partials.
>
> "Now the same piece is played with both melodic and harmonic scales stretched logarithmically in such a way that the octave ratio is 2.1 to l.
>
> "In the next presentation you hear the same piece with only the melodic scale stretched.
>
> "In the final presentation only the partials of each voice are stretched."

References

- E.Cohen (1984), "Some effects of inharmonic partials on interval perception," Music Perc. 1, 323-349.
- H.F.I. von Helmholtz (1877), *On the Sensations of Tone*, 4th ed., Transl. A.J.Ellis, (Dover, New York, 1954).
- M.V.Mathews and J.R.Pierce (1980), "Harmony and nonharmonic partials," J. Acoust. Soc. Am. 68, 1252-1257.
- F.H.Slaymaker (1970), "Chords from tones having stretched partials," J. Acoust. Soc. Am. 47, 1569-1571.

32. Primary and Secondary Beats (Starts at 1:05:55, length is 1:38)

If two pure tones have slightly different frequencies f1 and f2 = f1 + Δf, the phase difference j2 – j1 changes continuously with time. The amplitude of the resultant tone varies between Al + A2 and Al – A2, where Al and A2 are the individual amplitudes. These slow periodic variations in amplitude at frequency Δf are called beats, or perhaps we should say primary beats, to distinguish them from second–order beats, that will be described in the next paragraph. Beats are easily heard when Δf is less than 10 Hz, and may be perceived up to about 15 Hz.

A sensation of beats also occurs when the frequencies of two tones f1 and f2 are nearly, but not quite, in a simple ratio. If f2 = 2f1 + δ (mistuned octave), beats are heard at a frequency d. In general, when f2 = (n/m)f1 +δ, mδ beats occur each second. These are called second–order beats or beats of mistuned consonances, because the relationship f2 = (n/m) f1, where n and m are integers, defines consonant musical intervals, such as a perfect fifth (3/2), a perfect fourth (4/3), a major third (5/4), etc.

Primary beats can be easily understood as an example of linear superposition in the ear. Second–order beats between pure tones are not quite so easy to explain, however. Helmholtz (1877) adopted an explanation based on combination tones (Demonstration 34), but an explanation by means of aural harmonics was favored by others, including Wegel and Lane (1924). This theory, which explains second–order beats as resulting from primary beats between aural harmonics of f1 and f2, predicts the correct frequency m f2 – n f1, but cannot explain why the aural harmonics themselves are not heard (Lawrence and Yantis, 1957).

An explanation which does not require nonlinear distortion in the ear is favored by Plomp (1966) and others. According to this theory, the ear recognizes periodic variations in waveform, probably as a periodicity in nerve impulses evoked when the displacement of the basilar membrane exceeds a critical value. This implies that simple tones can interfere over much larger frequency differences than the critical bandwidth, and also that the ear can detect changing phase (even though it is a poor detector of phase in the steady state).

Beats of mistuned consonances have long been used by piano tuners, for example, to tune fifths, fourths, and even octaves on the piano. Violinists also make use of them in tuning their instruments. In the case of musical tones, however, primary beats between harmonics occur at the same rate as second–order beats, and the two types of beats cannot be distinguished.

In the first example, pure tones having frequencies of 1000 and 1004 Hz are presented together, giving rise to primary beats at a 4–Hz rate.

In the next example, tones with frequencies of 2004 Hz, 1502 Hz, and 1334.67 Hz are combined with a 1000–Hz tone to give secondary beats at a 4–Hz rate (n/m = 2/1, 3/2, and 4/3, respectively).

It is instructive to compare the apparent strengths of the beats in each case.

Commentary

> "Two tones having frequencies of 1000 and 1004 Hz are presented separately and then together. The sequence is presented twice.
>
> "Pairs of pure tones are presented having intervals slightly greater than an octave, a fifth and a fourth, respectively. The mistunings are such that the beat frequency is always 4 Hz when the tones are played together."

References

- H.L.F. von Helmholtz (1877), *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th ed. Transl. A.J.Ellis (Dover, New York, 1954).
- M.Lawrence and P.A.Yantis (1957), "In support of an 'inadequate' method for detecting 'fictitious' aural harmonics," J. Acoust. Soc. Am. 29, 750-51.
- R.Plomp (1966), *Experiments on Tone Perception*, (Inst. for Perception RVO– TNO, Soesterberg, The Netherlands).
- T.D.Rossing (1982), *The Science of Sound* (Addison-Wesley, Reading, MA). Chap. 8
- R.L.Wegel and C.E.Lane (1924), "The auditory masking of one tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. 23, 266-85.

33. Distortion (Starts at 1:07:44, length is 2:24)

This demonstration illustrates some audible effects of distortion external to the auditory system. These effects are of interest, not only because distortion commonly occurs in sound recording and reproducing systems, but because distortion is an important topic in auditory theory. This demonstration replicates one presented by W.M.Hartmann in the "Harvard tapes" (Auditory Demonstration Tapes, Harvard University, 1978). Both harmonic and intermodulation distortion are illustrated.

Our first example presents a 440–Hz sinewave tone, distorted by a symmetrical compressor.

A symmetrical compressor has an input–output relation such as that shown at the left. The important property is that the function describing the relation between input and output is an odd function—that is, f(x) is equal to –f(–x). Because of the symmetry, only odd harmonics of the original sinewave are present in the output. A simple example of a symmetrical compressor would be a limiter. In this demonstration, the distorted tone alternates with its 3rd harmonic (which serves as a "pointer").

Next the 440–Hz tone is distorted asymmetrically by a half–wave rectifier, which generates strong even–numbered harmonics. The distorted tone alternates with its 2nd harmonic.

When two pure tones (sinusoids) are present simultaneously, distortion produces not only harmonics of each tone but also tones with frequencies $n f_1 – m f_2$, where m and n are integers. The prominent cubic difference tone ($2f_1 – f_2$) which occurs when tones of 700 and 1000 Hz are distorted by a symmetrical compressor alternates with a 400–Hz pointer in the third example.

As a general rule the ear is rather insensitive to the relative phase angles between low–order harmonics of a complex tone. Distortion, however, especially if present in the right amount, can produce noticeable changes in the perceived quality of a complex tone when phase angles are changed. This is shown in the last demonstration in which the phase angle between a 440–Hz fundamental and its 880–Hz second harmonic is varied first without distortion and with the complex fed through a square–law device.

Commentary

> "First you hear a 440–Hz sinusoidal tone distorted by a symmetrical compressor. It alternates with its 3rd harmonic.

"Next the 440–Hz tone is distorted asymmetrically by a half–wave rectifier. The distorted tone alternates with its 2nd harmonic.

"Now two tones of 700 and 1000 Hz distorted by a symmetrical compressor. These tones alternate with a 400–Hz pointer to the cubic difference tone.

"You will hear a 440–Hz pure tone plus its second harmonic added with a phase varying from minus so to plus so degrees. This is followed by the same tones distorted through a square–law device."

References

- J.L.Goldstein (1967), "Auditory nonlinearity," J. Acoust. Soc. Am. 41, 676-89.
- J.L.Hall (1972a), "Auditory distortion products $f_2 - f_1$ and $2f_1 - f_2$", J.Acoust. Soc. Am. 51, 1863-71.
- J.L.Hall (1972b), "Monaural phase effect: Cancellation and reinforcement of distortion products $f_2 - f_1$ and $2f_1 - f_2$," J. Acoust. Soc. Am. 51, 1872-81.

34. Aural Combination Tones (Starts at 1:10:18, length is 1:40)

When two or more tones are presented simultaneously, various nonlinear processes in the ear produce combination tones similar to the distortion products heard in Demonstration 33—and many more. The most important combination tones are the difference tones of various orders, with frequencies $f_2 - f_1$ and $f_1 - n(f_2 - f_1)$, where n is an integer.

Two prominent difference tones are the quadratic difference tone $f_2 - f_1$, (sometimes referred to simply as "the difference tone") and the cubic difference tone $2f_1 - f_2$. If f_1, remains constant (at 1000 Hz in this demonstration) while f_2 increases, the quadratic difference tone moves upward with f_2 while the cubic difference tone moves in the opposite direction. At low levels (approximately 50 dB), they can be heard from about f_2/f_1, = 1.2 to 1.4 (solid line in the figure below), but at 80 dB they are audible over nearly an entire octave (f_2/f_1 = 1 to 2 dashed line in the figure below). In this case, quadratic and cubic difference tones cross over at f_2/f_1 = 1.5. They are shown on a musical staff in the right half of the figure.

In this demonstration, which follows No. 19 on the "Harvard tapes," tones with frequencies of 1000 and 1200 Hz are first presented. When an 804–Hz tone is added, it beats with the 800–Hz (quadratic) difference tone.

Next, the frequency of the upper tone is slowly increased from f_2 = 1200 to 1600 Hz and back again. You should hear the cubic difference tone moving opposite to f_2 soon joined by the quadratic difference tone, which first becomes audible at a low pitch and moves in the same direction as f_2. They should cross over when f_2 = 1500 Hz.

Over part of the frequency range, the quartic difference tone $3f_1 - f_2$ may be audible.

Commentary

"In this demonstration two tones of 1000 and 1200 Hz are presented. When an 804–Hz probe tone is added, it beats with the 800–Hz aural combination tone.

"Now the frequency of the upper tone is slowly increased from 1200 to 1600 Hz and then back again."

References

- J.L.Goldstein (1967), "Auditory nonlinearity," J. Acoust. Soc. Am. 41, 676-89.
- J.L.Hall (1972), "Auditory distortion products $f_2 - f_1$ and $2f_1 - f_2$," J. Acoust. Soc. Am. 51, 1863-71.
- R.Plomp (1965), "Detectability threshold for combination tones," J. Acoust. Soc. Am. 37, 1110-23.
- T.D.Rossing (1982), *The Science of Sound* (Addison-Wesley, Reading, MA). Chap. 8.

- B.Scharf and A.J.M.Houtsma (1986), "Audition 11: Loudness, pitch, localization, aural distortion, pathology," in *Handbook of Perception and Human Performance*, ed. L.Kaufman, K.Boff and J.Thomas (Wiley, New York).
- G.F.Smoorenburg (1972a), "Audibility region of combination tones," J. Acoust. Soc. Am. 52, 603-14.
- G.F.Smoorenburg (1972b), "Combination tones and their origin," J. Acoust. Soc. Am. 52, 615-632.

35. Effect of Echoes (Starts at 1:12:06, length is 1:51)

This so–called "ghoulies and ghosties" demonstration (No.2 on the "Harvard tapes") has become somewhat of a classic, and so it is reproduced here exactly as it was presented there. The reader is Dr. Sanford Fidell.

An important property of sound in practically all enclosed space is that reflections occur from the walls, ceiling, and floor. For a typical living space, 50 to 90 percent of the energy is reflected at the borders. These reflections are heard as echoes if sufficient time elapses between the initial sound and the reflected sound. Since sound travels about a foot per millisecond, delays between the initial and secondary sound will be of the order of 10 to 20 ms for a modest room. Practically no one reports hearing echoes in typical small classrooms when a transient sound is initiated by a snap of the fingers. The echoes are not heard, although the reflected sound may arrive as much as 30 to 50 ms later. This demonstration is designed to make the point that these echoes do exist and are appreciable in size. Our hearing mechanism somehow manages to suppress the later–arriving reflections, and they are simply not noticed.

The demonstration makes these reflections evident, however, by playing the recorded sound backward in time. The transient sound is the blow of a hammer on a brick, the more sustained sound is the narration of an old Scottish prayer. Three different acoustic environments are used, an anechoic (echoless) room, a typical conference room, similar acoustically to many living rooms, and finally a highly reverberant room with cement floor, hard plaster walls and ceiling. Little reverberation is apparent in any of the rooms when the recording is played forward, but the reversed playback makes the echoes evident in the environment where they do occur.

Note that changes in the quality of the voice are evident as one changes rooms even when the recording is played forward. These changes in quality are caused by differences in amount and duration of the reflections occurring in these different environments. The reflections are not heard as echoes, however, but as subtle, and difficult to describe, changes in voice quality. All recordings were made with the speaker's mouth about 0.3 meters from the microphone.

Commentary

> "First in an anechoic room, then in a conference room, and finally in a very reverberant space, you will hear a hammer striking a brick followed by an old Scottish prayer. Playing these sounds backwards focuses our attention on the echoes that occur."

References

- L.Cremer, H.A.Muller, and T.J.Schultz (1982), *Principles and Applications of Room Acoustics*, Vol.l (Applied Science, London).
- H. Kutruff (1979), *Room Acoustics*, 2nd ed. (Applied Science, London).
- V.M.A.Peutz (1971), "Articulatory loss of constants as a criterion for speech transmission in a room," J.Audio Eng. Soc. 19, 915-19.
- M.R.Schroeder (1980), "Acoustics in human communication: room acoustics, music, and speech," J. Acoust. Soc. Am. 68, 22-28.