© S. Hirzel Verlag · EAA
ACUSTICA · acta acustica
Vol. 82 (1996) 346 – 355

Auditory Localization Demonstrations

R. O. Duda
Department of Electrical Engineering, San Jose State University, San Jose, CA 95192

Summary
This collection of digitized sound examples was assembled for *Acta Acustica* to illustrate some interesting and important spatial hearing phenomena. The examples are grouped as follows:

1.	1 – 2	Calibration signals	(sinusoids)
2.	3	Monotic, Diotic and Dichotic Signals	(percussion)
3.	4	Binaural Recordings and Externalization	(clarinet)
4.	5 – 9	Monaural versus Stereo versus Binaural	(jet plane)
5.	10 – 13	Binaural Speech Separation	(speech)
6.	14	Reverberant and Anechoic Environments	(speech)
7.	15 – 17	Localization in Reverberant Environments	(tones)
8.	18 – 21	The Clifton Effect	(clicks)
9.	22 – 25	The Clifton Effect with Speech	(speech)
10.	26 – 29	The Clifton Effect with a Moving Echo	(clicks)
11.	30 – 33	The Clifton Effect with a Moving Echo	(speech)
12.	34 – 35	The Clifton Effect with Paired Echoes	(clicks)
13.	36 – 43	Binaural Adaptation	(various)
14.	44 – 48	Elevation Effects	(bells)

A listing of the announcements on the recording, keyed to CD-indices, is given at the end of this document.
Except for the Franssen-effect examples in Section 7, the recordings are meant to be heard through headphones. In listening to these examples, it is best to close one's eyes and to try to avoid head motion, paying particular attention to perceived spatial locations of the sounds. Reading the text for each section before listening may also be helpful, since what one should attend to is not always obvious. Finally, it should be noted that while some of the phenomena are well known and easily replicable, others are not well understood and are not perceived in the same way by all listeners. This illustrates why spatial hearing continues to be such an active and interesting area of psychoacoustic research.

PACS no. 43.66.Pn, 43.66.Qp, 43.66 Rq

1. Calibration Signals

A majority of the examples involve binaural hearing, and binaural demonstrations do not work well if the two channels are seriously unbalanced. Thus, the following two test signals are included to allow listeners to adjust their equipment.

a. 1000-Hz Tone Example 1

This test tone is at maximum amplitude on both channels. Use it to balance the channels for loudness, and to set the overall sound level below the point where it is uncomfortable or where distortion is audible.

b. Frequency Sweep — Left, Right Example 2

Use this swept tone to identify the left and right channels, and to check for freedom from resonant "buzzes" and for reasonable left/right balance throughout the audible spectrum. If the balance shifts significantly with frequency, note that this can be due to your hearing as well as to your equipment.

2. Monotic, Diotic and Dichotic Signals

This example illustrates three basically different headphone listening situations:

Monotic: signal in one headphone only
Diotic: same signal in each headphone
Dichotic: different signals in each headphone

In addition, it also illustrates how sounds heard through headphones can give a radically different impression than sounds heard through loudspeakers.

The example consists of four different presentations of percussion sounds[1] presented in sequence:

1. Sound in left channel only Example 3
2. Sound in right channel only
3. Same sound in both channels
4. Stereo sound in both channels

Heard through properly phased and balanced loudspeakers, these four excerpts sound rather similar, although the

Received 15 December 1994,
accepted 9 March 1995.

---

[1] Source: Ivan Lins and Victor Martins, "Lua Soberana," arranged and produced by Sergio Mendes, in *Brasileiro* (Elektra 61315-2).

stereo version clearly has some instruments on the left, some on the right, and the rest more or less in the center. This reflects the conventional way that stereo recordings are made by mixing multitrack recordings, where the recording producer assigns particular sound sources to desired spatial locations by the percentage of their signals used on the left and the right, usually without regard for phase or timing.

The same sounds heard through headphones produce quite different perceptions. The two monotic presentations are rather oppressive and unpleasant to listen to except at low volume levels. This is probably because the auditory system interprets such an unbalanced situation as corresponding to a sound source that is very close to one ear, which is a threatening situation.

The diotic presentation provides relief. The sound can be tolerated and even enjoyed at a much higher volume level (which can be injurious to one's hearing). Many listeners locate the sounds within their head, often distinctly above the interaural axis. However, subjective impressions of localization are notoriously variable from person to person.

The dichotic presentation is even more enjoyable. The sounds "open out" and assume different spatial locations. However, although there is a nice left/right spread of sound, the sounds usually seem to be located either within or very close to the head. It is often said that the stereo presentation over headphones produces lateralization rather than localization, and that the sounds are not externalized.[2]

3. Binaural Recordings and Externalization

This recording of a clarinet and drums [3] illustrates how the apparently small differences in the signals captured by the left and right ears provide localization information to the auditory system. The binaural recording was made using a Neumann KU81i artificial head (Kunstkopf), with the clarinet on the right and the drums slightly to the left of center.

There are three versions in sequence: Example 4

1. Left-ear signal on both channels
2. Right-ear signal on both channels
3. Binaural

In the first two cases — as usually happens with diotic signals — the sound is not externalized. The clarinet is a bit softer and duller sounding on the left channel than on the right channel, but the differences do not seem to be great. However, the binaural presentation is strikingly different. The clarinet is not only localized on the right (and perhaps a bit elevated), but also appears to be some distance away, and is clearly externalized. This ability to create an externalized sound image is one of the hallmarks of good binaural recordings.

4. Monaural versus Stereo versus Binaural

In this example, two Bruel and Kjaer microphones were used in a binaural recording of a jet plane[4] flying overhead from right to left. Several variations of that recording are presented.

a. Monaural. (Example 5) The left and right channels were simply mixed to produce a diotic presentation. As usual, the sound is not externalized, although some listeners experience a sense of motion, probably due to familiarity of the source and the obvious Doppler shift.

b. Stereo. (Example 6) The monaural recording was "cross-faded" to produce a stereo recording. That is, the right-channel signal was created by reducing the amplitude of the monaural signal starting at the time when the change in pitch due to Doppler shift was greatest. Similarly, the left-channel signal was created by increasing the amplitude of the monaural signal, with full amplitude reached at the same time that the right channel starts to fade out. This gives a fairly convincing illusion when heard over loudspeakers. When heard over headphones, however, the sound is not externalized, and the aircraft appears to fly through one's head from the right ear to the left ear.

c. Binaural — 44.1 kHz. (Example 7) Here the image of the jet is convincingly externalized, and the final sounds seem quite distant. Some listeners obtain a good sense of the aircraft passing overhead, while others feel that it is passing behind them. These differences in perception illustrate the fact that vertical localization depends on pinna effects, which vary significantly from person to person. A standard artificial head may produce a very good (or at least satisfactory) image for some listeners, and a weak or very confused image for others. This subject-to-subject variability — together with head-motion problems — is a major reason why binaural recordings have not been more widely accepted as commercial products.

d. Binaural — 22.05 kHz. (Example 8) To illustrate the importance of bandwidth, the original binaural recording was down-sampled by a factor of two to cutoff everything above 11 kHz. Those listeners who localized the full-bandwidth source overhead usually find that this sound image seems to be lower.

e. Binaural — 11.025 kHz. (Example 9) A second down-sampling was used to remove all high-frequency content above 5.5 kHz. Although externalization is still good, elevation effects are essentially lost. This indicates that of the three spherical coordinates — azimuth, elevation and range — it is elevation that requires the greatest bandwidth.

---

[2] There is a large literature on both the psychology and the technology of spatial hearing. The references cited at the end of this paper merely provide a few entry points to this literature.

[3] Musical source: John Barnes, "Sileypud," performed by New Haranni Poison Mixers and John Barnes, in *Stakkato Spezial* (AUDIO-CD 101020, AUDIO, Vereinigte Motor-Verlage GmbH & Co. KG).

[4] Sound source: Track 29, "Flugzeugstart: Boeing 737," in *Stakkato Spezial* (AUDIO-CD 101020, AUDIO, Vereinigte Motor-Verlage GmbH & Co. KG).

The cues for range are not as well understood as the cues for azimuth and elevation. The absolute intensity and spectral changes due to molecular absorption are major cues when the source is familiar, as is the case in this example. Neither of these cues is significantly weakened by loss of high- frequency information. However, these cues are also present in the monaural recording, which is not externalized. This illustrates the difference between estimating the range to a sound source and perceiving the location of a sound image. Interaural intensity differences are important for very close sources, and the ratio of direct to reverberant sound is important for distant sources. Head motion also plays a role, and provides a partial explanation for front/back reversals and the fact that people listening to binaural recordings frequently mention that the distance to sources located directly ahead seems compressed relative to the distance to sources at the side.

5. Binaural Hearing and the Cocktail-Party Problem

In his 1957 book entitled *On Human Communication*, Colin Cherry made the following observation: "One of our most important faculties is our ability to listen to, and follow, one speaker in the presence of others. This is such a common experience that we may take it for granted; we may call it 'the cocktail party problem.' No machine has been constructed to do just this, to filter out one conversation from a number jumbled together ...". Despite some significant recent progress, we still do not have a machine that can solve Cherry's problem.

In fact, people with hearing loss (particularly with loss of hearing in one ear) find it extremely difficult to comprehend speech in the simultaneous presence of other, similar interfering sounds. Research on auditory scene analysis has shown that spatial localization is only one of the many characteristics of different sound sources that we exploit in solving the cocktail party problem. However, it is definitely easier to separate sources heard binaurally than sources heard monaurally.

Two sequences are given to illustrate this effect. In the first, two different people are "conversing." In the second, two recordings of the same person were mixed. While there are ample cues for separation even in the monaural recordings, the binaural presentations make separation more effortless.

1. Different talkers
   a) Monaural presentation — Example 10
      Talker 1
      Talker 2
      Mixture
   b) Binaural presentation — Example 11
      Talker 1
      Talker 2
      Mixture
2. Same talker
   a) Monaural presentation — Example 12
      Sentence 1
      Sentence 2
      Mixture
   b) Binaural presentation — Example 13
      Sentence 1
      Sentence 2
      Mixture

6. Reverberant and Anechoic Environments

Even when there is only one sound source active, the auditory system must cope with the multipath problems — echoes and reverberation that are almost always present but are rarely consciously noticed.

A reflection from only one surface (such as the ground) is sufficient to play havoc with the waveforms arriving at the ear. In the time domain, these reflections show up more or less as smaller, delayed replicas that are added to the original sound. In the frequency domain, they introduce "comb-filter" effects into frequency responses; at some frequencies, the secondary waves are reinforcing and produce response peaks, while at others they are interfering and produce response dips. This can introduce a "pitchy" quality called repetition pitch in otherwise pitch-free sounds. Given the degree to which reflections from environmental surfaces can distort both waveforms and frequency responses, it is surprising that we can hear anything intelligible in enclosed spaces.

To reduce the effects of room reflections on acoustic and auditory measurements, researchers in the 1930's built anechoic chambers — rooms whose sides, floor and ceiling are lined with deep wedges of sound absorbing materials. Example 14 is a binaural recording made with a KEMAR[5] artificial head located in an anechoic chamber. When the person speaking moves from the outside of the chamber to the inside, the differences in the "liveness" and "quality" of the sound are quite apparent. However, the many reflections that exist in the reverberant environment become apparent only when the room is large enough to hear actual echoes.

7. Localization in Reverberant Environments

7.1. The Franssen Effect

When wide-band sounds are heard in enclosed, reverberant spaces, reflections (echoes) usually do not prevent the listener from locating the source. In general, sounds are localized on the basis of the waves that arrive first, which take precedence over the subsequent reflections. This is variously referred to as the precedence effect, the Haas effect, or the law of the first wavefront. The Franssen effect is a well known, dramatic illustration of the power of the precedence effect. However, the

---

[5] KEMAR — which stands for Knowles Electronics Manikin for Acoustic Research — is a standard and anatomically faithful artificial head used in hearing-aid development.

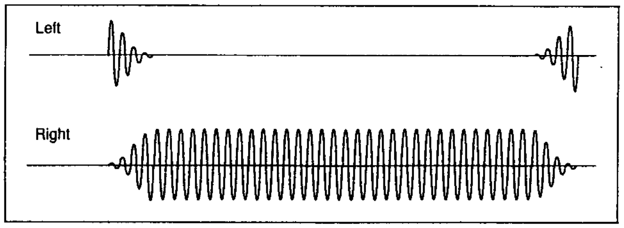

Figure 1. Shape of the basic signals used in the Franssen-effect experiments.

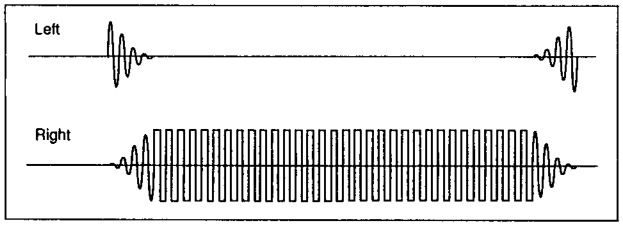

Figure 2. Clipped signals used in the Franssen-effect experiments.

test signals are intended for loudspeakers, and when the signals are heard through headphones, the perception is entirely different.

The basic signals used are 200-Hz sinusoids that last about 3 seconds. The left signal is windowed to have an exponential onset and offset, with a 100-ms time constant. The right signal is the complement of the left signal, and thus basically consists of an onset transient, followed 3 seconds later by an offset transient. (Example 15; see Figure 1) In listening to these signals, the levels must be kept low enough so that there is no clipping distortion, and the loudspeaker cabinets should be free from buzzes or rattles.

When the signals are sent to loudspeakers, listeners will first hear the onset transient from the left speaker, and will usually localize the sound there from then on, despite the fact that after about 400 ms essentially all of the energy is coming from the right speaker. The illusion is frequently maintained even when the listener moves close to each speaker in an active attempt to locate the source.

However, when the same signals are played through headphones, the effect is entirely different, and the listener hears the signals for what they really are.

7.2. The Effect of Clipping on the Franssen Effect.

The low-frequency sinusoid used in the Franssen effect is basically a narrow-band signal, which is what makes it so hard to localize. If it were not for the high-frequency energy in the onset, it would be difficult to locate such a sound source in a reverberant room. In particular, once the transient buildup has passed, sine waves set up very complicated standing wave patterns in rooms, making them much harder to locate than wide-band sounds.

It is easy to illustrate the effect of bandwidth by clipping the signals used in the Franssen-effect experiment. (Example 16; see Figure 2.) Such clipped signals are rich in harmonics, and the location of the source is as easy to determine with loudspeakers as with headphones.

In the demonstration sound, only the right channel is clipped. Notice the brief onset in the left speaker, followed by the long, clarinet-like clipped tone in the right speaker, terminated by the brief offset back in the left speaker. Much the same perception is experienced through headphones.

7.3. Localization of Sinusoids by Time Difference

If the abrupt onset is exploited by the auditory system to localize a sound, are onsets required for localization? The answer is a bit complicated.

In this experiment, the 200-Hz left and right signals are essentially constant, but are windowed to provide a very gradual onset and offset. In addition, the signal in the right channel is slowly advanced and then slowly delayed with respect to the left channel. The sequence (Example 17) is as follows:

0 to 2 sec	Both channels identical
2 to 3 sec	Gradual introduction of phase shift
3 to 5 sec	Right channel delayed by 0.65 ms
5 to 7 sec	Gradual reversal of phase shift
7 to 9 sec	Left channel delayed by 0.65 ms
9 to 10 sec	Gradual reversal of phase shift
10 to 12 sec	Both channels identical

The phase shifts are introduced so slowly that either channel heard alone sounds like a steady 200-Hz tone. However, when heard through headphones, the lateralizations implied by the above table are clearly heard: center, left, right, and center. Since the ability of the auditory system to detect such phase shifts in pure, low-frequency sine tones has been known since the 1930's, this result is hardly surprising.

Heard through loudspeakers in a reverberant room, however, the perception can be quite different. Exactly what this test signal sounds like depends upon such factors as the size and shape of the room, the absorption characteristics of the reflecting surfaces, and the locations of the loudspeakers. However, a common experience is that the location of the source in this example can be surprisingly ambiguous.

In talking about researchers who did the early work on sound localization, Mills remarked that "... Their efforts to measure the roles of interaural differences in time and intensity were hampered by the lack of equipment for producing well-controlled sounds and by their custom of experimenting in reverberant rooms. Better measurements of the localization of actual sources waited upon the means to generate pure tones and to present them in a space free from reflected sounds." However, natural sounds in everyday environments are quite different from pure tones heard over headphones or in anechoic chambers, and researchers are once again investigating complex sounds in more natural settings to gain a better understanding of auditory localization.

Thus, the conclusion is that although the auditory system certainly possesses the ability to localize narrow-band, low-frequency sounds, this ability is of limited use in reverberant rooms. In reverberant environments, the auditory system exploits the broad-band onsets and often ignores low- frequency phase shift. In general, our perceptual systems use multiple sources of information, and seem to exploit the most reliable ones in any particular situation.

8. The Clifton Effect: Click Stimulus

The Franssen effect and many other experiments show that our auditory systems can be so effective at suppressing reflected sounds that we are usually not consciously aware of their presence. However, this describes a steady-state situation that exists when the listener has adjusted to the acoustic environment, and it leaves questions about the dynamics of the process unanswered.

In 1987, Clifton showed that the precedence effect can break down, and that it can take a remarkably long time (up to a few seconds) for it to become re-established. This showed that echo suppression does not always take place immediately, and shows that models that employ fast-acting mechanisms in which sudden onsets quickly inhibit signals that follow are incomplete. Instead, it suggests that echo suppression also entails central processes that require recognition of the stability of the source/echo pattern, and might be involved in such activities as sound-source formation and acoustic-environment modeling. In Clifton's basic experiment, the listener is seated between two loudspeakers in an anechoic chamber and listens to short-duration clicks. A click played through one speaker (the source) is followed T seconds later by a click from the other speaker (the echo). Under ordinary conditions, if this pattern is heard repeatedly, and if T is less than roughly 5 msec, the echo is suppressed, and the listener hears only the source. This is consistent with the precedence effect. In Clifton's experiment, the click pairs are presented repeatedly a number of times, and then the signals to the two speakers are suddenly switched. The result is surprising. For a brief time (usually lasting four or five clicks) the listener hears both clicks. However, the "echo" then fades away, and finally only the source click is heard.

Although anechoic chambers are not widely available, a similar experiment can be simulated using headphones. In the present example, the basic stimulus is a rectangular pulse (click) of 160-μsec duration (7 samples at 44.1 kHz). To simulate the effects of head diffraction, the signal for the contralateral ear is computed by delaying the click by 0.65ms and filtering it with a low-pass filter having a 1500-Hz cutoff frequency. If this pulse is heard binaurally by itself a single time, it is immediately and confidently lateralized. (Example 18).

A simulated echo is constructed by interchanging the right and left signals and introducing a 5 ms delay, and the click and its echo are mixed to form a click/echo pair (see Figure 3). If this click/echo pair is heard only once, one is aware of a click in each ear, and may be able to tell that the echo is being heard slightly after the click. Upon repeated hearing, the echo seems to fade away (Example 19). In itself, this shows that the precedence effect is not instantaneous, but develops dynamically. The final perception is not the same as that of the isolated click. Some listeners describe it as a "fuller sounding" pulse located more or less where the initial click is, as might be heard in a reverberant environment. The effects of the echo remain audible, but as a minor distraction.

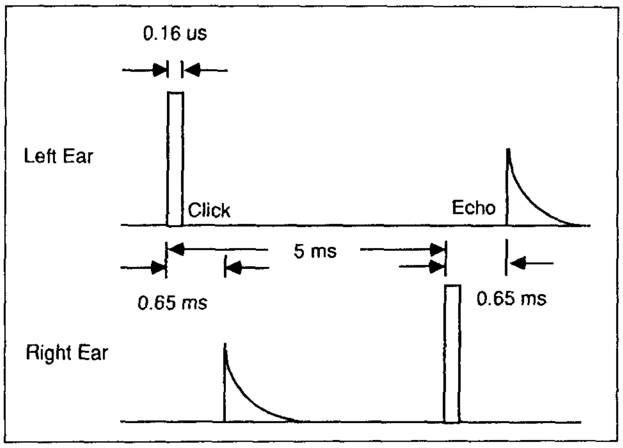

Figure 3. Click/echo pairs used in the experiments.

Finally, if the right and left signals are switched after adaptation, there is a brief period (3 or 4 click pairs) during which both sounds are heard. After that, the sound from the initial location fades out, and the sound is more or less localized on the final location. This is what Clifton refers to as the "breakdown of the precedence effect" (Examples 20 and 21). Two versions are given, since Clifton found that some people adapt more quickly when the leading sound is on the right than when it is on the left.

9. The Clifton Effect: Speech Stimulus

In this experiment a speech sound was substituted for the clicks. The word "talking" was recorded binaurally using KEMAR in the San Jose State University anechoic chamber. An "echo" from the opposite side was created by interchanging the left and right signals. A signal/echo pair was produced by mixing the signal with the "echo" delayed by 10 ms, which is twice the delay used for the clicks. (Example 22)

The results are much the same as with the clicks, except that adaptation occurs with fewer repetitions (although one can argue that there are many "events" per repetition).

Specifically, if the signal/echo pair is heard only once, one is aware of two talkers in each ear, with the echo being heard slightly after the signal. Again, upon repeated hearing, the echo seems to fade away (and perhaps move downward and outward), though the perception is not the same as that of the isolated word. Some listeners describe it as a "brighter" word located on the side of the initial speech but somewhat further away, somewhat further to the side, and heard in a reverberant environment. The echo remains audible, especially its higher frequencies, but as a minor distraction. (Example 23)

Again, if the right and left signals are switched after adaptation, there is a brief period (2 or 3 instances) during which both sounds are heard. (This is Clifton's "breakdown of the precedence effect.") After that, the sound from the initial location fades out, and the sound is more or less localized on the final location. (Examples 24 and 25) We conclude that the Clifton effect is not a special phenomenon associated with unusual impulsive sounds, but is a common characteristic of the precedence effect. That makes it rather surprising that we are not aware of these effects from ordinary, day-to-day experience.

10. The Clifton Effect: Centered Click and Moving Echo

One problem with the classic Clifton experiments is that both the source and the echo move. When the right/left switch occurs, the auditory system suddenly confronts a new source on the opposite side. Thus, what we may have is a "source formation effect," not a breakdown in the precedence effect. In particular, suppose that a new source suddenly appeared on the opposite side of an old source that becomes silent at the same instant. One might not be surprised that the auditory system interpreted the echo of the new source as a continuation of the existence of the old source. That is, the time constants may have to do with the speed with which the auditory system can commit to the existence of a new source, and may not be directly related to echo suppression.

To investigate this, we made a slight modification to the earlier experiment. The source always consisted of two simultaneous clicks, and thus appeared to be localized in the center of the head. The echo was delayed by 5 ms and was a pair of clicks that included simulated head shadow.

The click and its echo were mixed to form a click/echo pair. If this click/echo pair is heard only once, one is aware of one click in the center and a second at one ear, with the echo perhaps being heard a very short time after the click. (Example 26) Again, upon repeated hearing, the echo seems to fade away, though the perception is not the same as that of the isolated click. Some listeners describe it as a "fuller sounding" pulse located more or less where the initial click is, as heard in a reverberant environment. The echo remains barely audible, and probably would not be noticed by most listeners. (Example 27)

If the right and left signals are switched after adaptation, a dramatic change once again occurs. There is a brief period (3 or 4 click pairs) during which the echo (now on the other side) jumps into prominence. Then the echo fades out as before, and the sound is more or less localized in the center again. (Examples 28 and 29)

This experiment makes it clear that the auditory system is by no means just "shutting down" following the first click, but is open to the appearance of the new echo. Given the long response time, one concludes that the suppression of the perception of the echo must be happening at centers that are central rather than peripheral.

11. The Clifton Effect: Centered Speech and a Moving Echo

To complete the pattern, this experiment substituted a speech sound for the clicks in the previous experiment. The signals were derived from the binaural recording of the word "talking." The left channel was copied into the right channel to produce a pseudo centered binaural image. The original "talking" recording was used for the "echo." A signal/echo pair was produced by mixing the signal with the "echo" delayed by 10 ms. (Example 30)

The results are more or less as expected. Specifically, if the signal/echo pair is heard only once, one is aware of two talkers, one in the center and one on the left (although it also sounds like a spatially extended source). Upon repeated hearing, much of the echo seems to fade away, though the perception is not the same as that of the isolated word. Some listeners describe it as a "brighter" word located in the center and heard in a reverberant environment. The echo remains audible, especially its higher frequencies, but as a somewhat noisy distraction. (Example 31)

Again, when the right and left signals are switched after adaptation, there is a brief period (2 or 3 instances) during which the new echo jumps into prominence. After that, the new echo fades out, and the sound is more or less localized on the final location. (Examples 32 and 33)

12. The Clifton Effect: Paired Echoes

Clifton's original experiment can be thought of as modeling a situation in which a click source is initially on the listener's left, a reflecting wall is on the right, and after some time the source and the wall instantly change places. The centered-click variation can be thought of as modeling a situation in which the click source is directly in front of the listener, and there is a reflecting wall on the listener's left that suddenly disappears and is replaced by a reflecting wall on the right. An interesting further variation of Clifton's experiment (originally suggested by Richard F. Lyon of Apple Computer, Inc.) employs two echoes. It can be thought of as the centered-click variation, but where the left reflecting wall does not disappear when the right reflecting wall appears. (In the example given in this section, no multiple reflections are simulated.)

To be more specific, the source always consisted of two simultaneous clicks, and thus appeared to be localized in the center of the head. Both the left and right echoes were delayed by 5 ms and were identical except for left/right reversal. (Example 34) The key feature of this variation is that at the end, when the click and both echoes are present, the right-ear and left-ear signals are identical. Thus, it is not surprising that the final perception is of a centered click with no particularly reverberant quality.

As with the previous centered-click example, upon repeated hearing, the initial echo seems to fade away. When the right echo is introduced after adaptation, one might expect to be suddenly aware of the right echo, and then to have its presence slowly fade away. Instead, however, the final perception of a centered click seems to develop immediately, with no noticeable adaptation time. (Examples 35)

This example shows that there need not be long adaptation times following a sudden change in the binaural stimulus, and that one's models of the acoustic environment can be rapidly updated. Indeed, in their studies of the decline in usefulness of interaural information after the initial onset of the signal, Hafter and his students concluded that a variety of changes in the stimulus can trigger a very rapid recovery from saturation. The next examples illustrate both slow and rapid forms of binaural adaptation.

13. Binaural Adaptation

Many important studies of binaural sound localization have been done by synthesizing stimuli that place ITD and IID cues in conflict. A classic example is the trading studies that reveal the relative importance of ITD and IID for sinusoids of different frequencies. The way that the auditory system reconciles inconsistent cues for more complex sounds is not completely understood, and seems to vary significantly from person to person. In this section, we present sequences of consistent and inconsistent stimuli involving clicks, music, and random noise. As with the Clifton effect, these examples involve adaptation and implicate central processes. They also illustrate some of the problems facing people who want to synthesize artificial binaural sounds to create specific impressions of spatial localization.

13.1. Clicks

In our Clifton-effect examples, we not only delayed the contralateral signal but also low-pass filtered it to provide consistent interaural time difference (ITD) and interaural intensity difference (IID) localization cues. These clicks sound rather dull, but they are immediately localized, and their location is stable. (Example 36)

If the low-pass filtering is omitted, the auditory system is presented with an inconsistent stimulus, the ITD information suggesting a sound source on the side, and the IID information suggesting a sound source in the center. Informal experiments reveal a variety of reactions to this stimulus. Many listeners report that clicks seem centered at first, but then drift to the side of the leading pulse. However, others report hearing two clicks initially, a persistent one on the left and a brighter or more reverberant one in the center that fades out. Furthermore, some listeners report hearing only a left or only a centered click, with no drift or adaptation. Some of these differences may be due to the fact that listeners can switch at will between an analytic and a synthetic mode of listening. However, the different reactions could also imply significant person-to-person variation in the preference of the auditory system for time or intensity cues. (Example 37)

If sequences of consistent and inconsistent clicks are alternated, the differences can be more directly compared. Our example consists of four repetitions of a sequence of 44 clicks, the first 4 being consistent and the subsequent 40 being inconsistent. (The time between clicks is always 0.2 sec. Example 38 for clicks on the left, and Example 39 for clicks on the right.) Those listeners who experience an adaptation to the inconsistent clicks typically find that each hearing of the consistent clicks "resets" things, and an adaptation to the conflicting IID/ITD cues must be repeated.

13.2. Music

While very short clicks are valuable stimuli for binaural hearing experiments, they are uncommon in nature. In this experiment, we use a guitar recording for the stimulus [4]. As in the previous case, we provide an example with the source localized on the left, and a mirror-image example with the source localized on the right. (As with Clifton's experiments, many people exhibit a right/left asymmetry in response to these stimuli.)

When the source is localized on the left, the right signal is delayed with respect to the left signal by 0.65 ms. The signal is divided into intervals. During the "consistent" intervals, the right signal is also low-pass-filtered (1,500- Hz cutoff frequency). During the "inconsistent" intervals, the only difference is time delay. The sequence is as follows:

Interval	Number of Musical Bars	State
1	4	Consistent
2	8	Inconsistent
3	4	Consistent
4	12	Inconsistent
5	6	Consistent
6	14	Inconsistent
7	1	Consistent

Example 40 is for the source on the left and Example 41 is for the source on the right.

A typical description of the perceptions for Example 40 is as follows. "During Interval 1, the source is immediately localized on the left, and does not drift. Call this S1. At the beginning of Interval 2, it is as if S1 stops and a new and brighter-sounding source S2 appears to the left of center and a bit elevated. S2 drifts somewhat to the left and drops a bit during the interval." However, there is as much variations in the perception of this sequence as occurs with the click stimuli.

---

[4] Sound source: M. Falla, "Spanish Dance No. 1" from *La Vida Brève*, performed by Christopher Parkening and David Brandon, in *Virtuoso Duets* (EMI/Angel CDC-7494062).
As it happens, there were actually two guitars in the recording. However, their monaural combination effectively provided a single source.

13.3. White Gaussian Noise

While musical sounds are more "natural" than clicks, they have strong periodic structure, which makes them harder to localize. (It is known, for example, that the echo threshold — the time delay required for a delayed sound to be perceived as a distinct echo — is much longer with music than with clicks.)

White gaussian noise provides another important test sound, and reveals yet another phenomenon. The experiment is as follows. Let $s(t)$ be a white gaussian noise signal, let $d(t)$ be a delayed version of $s(t)$, and let $f(t)$ be a low-pass-filtered version of $d(t)$. Thus, using our earlier terminology, the binaural pair (s, d) is inconsistent, whereas the binaural pair (s, f) is consistent. These sounds are presented in the sequence shown in table I, in which consistent and inconsistent pairs alternate.

Again, two versions are provided, one for the sound on the left (Example 42) and one for the sound on the right (Example 43).

As with the earlier examples, there is significant person-to- person variation in the response to this stimulus. The following is a description of my personal experience when the source is on the left.

- Interval 1 immediately sounds like a noise source S1 located on the left.
- At the start of Interval 2, a second, "brighter" source S2 jumps into prominence near the center, and although it does not disappear, S1 seems to drop in prominence. S2 slowly fuses with S1 into a single but brighter source. With Interval 3, the original source S1 is quickly re- established, with no apparent drift.
- At the start of Interval 4, S2 again appears. This time, however, S1 remains fairly prominent, and even at the end of Interval 4, both S1 and S2 seem to be individually present, with S1 duller and S2 brighter.
- Interval 5 exhibits S1 as usual.
- Interval 6 is more or less like Interval 4, although the separate existence of both S1 and S2 seems even better established. Interval 7 is physically the same as Interval 1, and gives rise to essentially the same perception.
- Interval 8 is physically the same as Interval 2, but it is perceived much more like Interval 6 — with both S1 and S2 retaining their separate identities.

This is clearly a more complex impression than was the case with clicks or the guitar. There seem to be at least two different kinds of adaptation involved:

1. Adjustment to the inconsistent cues under the assumption that there is only one sound source present.
2. Source formation decisions. Initially, a single source hypothesis is simpler, and S1 is allowed to disappear from consciousness at the start of Interval 2. However, after repeated exposure to S1 and S2, the evidence mounts for the existence of two sources, which are allowed to maintain their separate identities.

It is not clear why one hears only one source in some situations but two sources in others. However, the long exposure times that are required to form these perceptions implicates central rather than peripheral processes.

14. Elevation Effects

It is easy to perceive changes in azimuth in binaural recordings, but it is much harder to perceive changes in elevation. The basic problem is that the primary elevation cues are derived from the shape of the outer ears or pinnae, and there is very significant person-to-person variation in pinna shapes. Most people experience good elevation effects when listening to recordings made through their own pinnae. However, localization accuracy can be dramatically degraded when listening through other people's pinnae. (In addition, some people have such unfortunately shaped pinnae that they are not good at localization in elevation under any circumstances.)

The following examples used head-related transfer functions measured for the KEMAR manikin. If you do not perceive the sounds in the locations described, it is probably because your pinnae differ from KEMAR's.

The examples consist of four sequences in which the sound of a small bell was convolved with the KEMAR head-related impulse responses to simulate systematic motion around simple but precise circular trajectories. (Example 44)

Sequence 1. (Example 45) The bell moves in the horizontal plane from the left ear in 17 steps around the front to the right ear. This is usually readily perceived.

Sequence 2. (Example 46) The bell moves in the vertical median plane from below to in front to above to behind to below. Although one can hear the monaural spectral cues to elevation (a brightening of the sound overhead and a dulling below), most listeners find that the elevation perception is quite unconvincing.

Sequence 3. (Example 47) The bell moves in the vertical frontal plane from the left to above to the right to below and back to the left in 34 steps. Although not perfect, this is usually more convincing than the median-plane case.

Sequence 4. (Example 48) The bell is on the right and moves around a 40-degree-azimuth cone, going from below to ahead to above to behind to below. Some listeners find this to be fairly convincing, with reasonable front/back discrimination, but the bell sometimes seems to "skip around" rather than to move monotonically from position to position. Others find the entire sequence to be spatially confusing, or experience no elevation effects. This illustrates the need for customized head-related transfer functions if one wants to produce reliable elevation effects in auditory displays.

Acknowledgement

This work was supported by the National Science Foundation under NSF Grant No. IRI-9402246. I am indebted to Richard Lyon of Apple Computer, Inc., for his long-term support, advice and encouragement. Not being a professional

Table I.

Interval	Time Interval	Duration [s]	Left ear	Right ear	State
1	0.0 to 0.5	0.5	$s(t)$	$f(t)$	Consistent
2	0.5 to 10.5	10	$s(t)$	$d(t)$	Inconsistent
3	10.5 to 11.5	1	$s(t)$	$f(t)$	Consistent
4	11.5 to 21.5	10	$s(t)$	$d(t)$	Inconsistent
5	21.5 to 23.5	2	$s(t)$	$f(t)$	Consistent
6	23.5 to 33.5	10	$s(t)$	$d(t)$	Inconsistent
7	33.5 to 34.0	0.5	$s(t)$	$f(t)$	Consistent
8	34.0 to 44.0	10	$s(t)$	$d(t)$	Inconsistent

psychoacoustician, I also greatly appreciate the many helpful suggestions made by Frans Bilsen, William Hartmann, Johan Raatgever, Malcolm Slaney, and additional anonymous reviewers, whose advice I have tried to follow.

Appendix

Announcements for the Examples

1. One thousand Hertz.
2. Twenty Hertz to twenty kilo-Hertz swept frequencies.
   Sweep on the left channel.
   Sweep on the right channel.
   Both left and right.
3. Four presentations of percussion sounds.
   Left only.
   Right only.
   Left and right.
   And stereo.
4. Three presentations of a binaural recording.
   Left-ear signal on both channels,
   right-ear signal on both channels,
   binaural.
5. Five presentations of a jet aircraft recording. Number 1. Monaural.
6. Number 2. Stereo.
7. Number 3. Binaural, 44.1 kilo-Hertz sampling rate.
8. Number 4. Binaural, 22 kilo-Hertz sampling rate.
9. Number 5. Binaural, 11 kilo-Hertz sampling rate.
10. The Cocktail-Party Problem.
    A. Utterances from two different talkers.
    Number 1. Monaural presentations.
    Talker 1.
    Talker 2.
    Mixture.
11. Number 2. Binaural presentations.
    Talker 1.
    Talker 2.
    Mixture.
12. The Cocktail-Party Problem
    B. Two different utterances from the same talker.
    Number 1. Monaural presentations.
    Sentence 1.
    Sentence 2.
    Mixture.
13. Number 2. Binaural presentations.
    Sentence 1.
    Sentence 2.
    Mixture.
14. An example of sounds heard binaurally in a reverberant and an anechoic environment.
15. Localization in reverberant environments.
    A. The Franssen effect, repeated twice.
16. B. The effect of clipping on the Franssen effect, repeated twice.
17. C. Localization of sinusoids by time difference, repeated twice.
18. The Clifton Effect using a click as a stimulus.
    A single click on the left.
    A click on the left and its echo on the right.
19. Forty repetitions of the click/echo pair
20. The Clifton Effect. Left- and right-ear signals are reversed after forty clicks.
21. The Clifton effect with the source initially on the right.
22. The Clifton Effect using speech as a stimulus.
    The word "talking" spoken on the left.
    The word and its echo.
23. Twenty repetitions of the word/echo pair.
24. The Clifton Effect. Left- and right-ear signals are reversed after ten repetitions.
25. The Clifton Effect with the source initially on the right.
26. The Clifton Effect with a centered click and a moving echo.
    The source click.
    The echo on the left.
    The click/echo pair.
27. Forty repetitions of the click/echo pair.
28. The Clifton effect. Left- and right-ear signals are reversed after forty clicks.
29. The Clifton Effect with the echo initially on the right.
30. The Clifton Effect with centered speech and a moving echo.
    The source word.
    The echo.
    The word/echo pair.
31. Twenty repetitions of the word/echo pair.
32. The Clifton Effect. Left- and right-ear signals are reversed after ten repetitions.
33. The Clifton Effect with the echo initially on the right.

34. A variation of the Clifton Effect suggested by Dick Lyon.
First the centered click.
Next the left echo.
Next the right echo.
35. Forty repetitions of the click, left, and right echoes.
36. Binaural adaptation. Clicks with consistent interaural time and intensity cues.
37. Clicks with inconsistent interaural time and intensity cues.
38. Alternation of 4 consistent and 40 inconsistent clicks. Left side.
39. Alternation of 4 consistent and 40 inconsistent clicks. Right side.
40. A guitar recording with alternating consistent and inconsistent cues. Guitar on the left.
41. A guitar recording with alternating consistent and inconsistent cues. Guitar on the right.
42. A recording of Gaussian white noise with consistent and inconsistent cues. Noise on the left.
43. A recording of Gaussian white noise with consistent and inconsistent cues. Noise on the right.
44. Elevation effects. The bell sound used as a source, repeated twice.
45. Motion in the horizontal plane. Seventeen positions from left to right, repeated twice.
46. Motion in the median plane. Twenty positions from below to front to above to behind back to below, repeated twice.
47. Motion in the frontal plane. Thirty-four positions from left to above to right to below back to left, repeated twice.
48. Motion on the right around a 40-degree-azimuth cone. Twenty positions from below to in front to above to behind to below, repeated twice.

References

Bassett, J. G. and Eastmond, E. J. (1964). Echolocation: measurement of pitch versus distance for sounds reflected from a flat surface. *J. Acoust. Soc. Am.* **36**, 911–916.

Bilsen, F. A. (1970). Repetition pitch and its implications for hearing theory. *Acustica* **22**, 64–73.

Blauert, J. (1974). *Spatial Hearing*. MIT Press, Cambridge, MA.

Blauert, J. (in press). An introduction to binaural technology. Gilkey, R. and Anderson, T., editors, *Binaural and Spatial Hearing*, Lawrence Elbaum Associates, Hillsdale, NJ.

Bodden, M. (1993). Modeling human sound-source localization and the cocktail-party-effect. *Acta Acustica* **1**, 43–55.

Bregman, A. S. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, MA.

Brown, G. J. and Cooke, M. (1993). Physiologically motivated signal representations for computational auditory scene analysis. M. Cooke, S. B. and Crawford, M., editors, *Visual Representations of Speech Signals*, John Wiley and Sons, Chichester, England, 181–188.

Cherry, C. (1957). *On Human Communication*. The Technology Press of Massachusetts Institute of Technology and JohnWiley and Sons, New York, NY.

Clifton, R. K. and Freyman, R. L. (1989). Effect of click rate and delay on breakdown of the precedence effect. *Perception and Psychophysics* **46**, 139–145.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., and Wenzel, E. M. (1992). On the externalization of auditory images. *Presence* **1**, 251–257 (Spring issue).

Franssen, N. V. (1960). *Some considerations of the mechanism of directional hearing*. PhD thesis, Technische Hogeschool, Delft, Delft, The Netherlands.

Hafter, E. R. and Buell, T. N. (1990). Restarting the adapted binaural system. *J. Acoust. Soc. Am.* **88**, 806–812.

Hartmann, W. M. (1983). Localization of sound in rooms. *J. Acoust. Soc. Am.* **74**, 1380–1391.

Hartmann, W. M. and Rakerd, B. (1989). Localization of sound in rooms iv — the franssen effect. *J. Acoust. Soc. Am.* **86**, 1366–1373.

Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (1987). Auditory demonstrations. Compact disc, Institute for Perception Research with the support of the Acoustical Society of America and produced by the Philips Company (CD 1126-061), Eindhoven, The Netherlands. This recording of 39 sound demonstrations includes the classic "ghoulies and ghosties" example that makes the reverberation apparent by playing the sounds backwards. It also contains examples of important binaural phenomena not directly related to localization.

Kawaura, J., Suzuki, Y., Asano, F., and Sone, T. (1991). Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *J. Acoust. Soc. Japan* **(E)12**, 203–216.

Mershon, D. H. and Bowers, J. N. (1979). Absolute and relative cues for auditory perception of egocentric distance. *Perception* **8**, 311–332.

Middlebrooks, J. C. and Green, D. M. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.* **42**, 135–159.

Mills, W. (1972). Auditory localization. Tobias, J. V., editor, *Foundations of Modern Auditory Theory, Vol. II*, Academic Press, New York, NY, 303–348.

Schröder, M. R. (1980). Acoustics in human communication: room acoustics, music, and speech. *J. Acoust. Soc. Am.* **68**, 22–28.

Sunier, J. (1989a). Binaural overview: Ears where the mikes are. part i. *Audio* **73**, 75–84 (November issue).

Sunier, J. (1989b). Binaural overview: Ears where the mikes are. part ii. *Audio* **73**, 49–57 (December issue).

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (1949). The precedence effect in sound localization. *Am. J. Psych.* **62**, 315–336.

Wenzel, E. M. (1992). Localization on virtual acoustic displays. *Presence* **1**, 80–107 (Winter issue).

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* **94**, 111–123.

Wightman, F. L. and Kistler, D. J. (1989a). Headphone simulation of free-field listening. i: Stimulus synthesis. *J. Acoust. Soc. Am.* **85**, 858–867.

Wightman, F. L. and Kistler, D. J. (1989b). Headphone simulation of free-field listening. ii: Psychophysical validation. *J. Acoust. Soc. Am.* **85**, 868–878.

Yost, W. A. (1991). Auditory image perception and analysis: The basis for hearing. *Hearing Research* **56**, 8–18.