Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces

Durand R. Begault and Elizabeth M. Wenzel

August 1990

NASA
National Aeronautics and
Space Administration

Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces

Durand R. Begault and Elizabeth M. Wenzel
Ames Research Center, Moffett Field, California

August 1990

CONTENTS

TECHNIQUES AND APPLICATIONS FOR BINAURAL SOUND MANIPULATION IN HUMAN-MACHINE INTERFACES

Durand R. Begault and Elizabeth M. Wenzel

Ames Research Center

SUMMARY

The implementation of binaural sound to speech and auditory sound cues ("auditory icons") is addressed from both an applications and technical standpoint. Techniques overviewed include processing by means of filtering with head-related transfer functions. Application to advanced cockpit human interface systems is discussed, although the techniques are extendable to any human-machine interface. Research issues pertaining to three-dimensional sound displays under investigation at the Aerospace Human Factors Division at NASA Ames Research Center are described.

INTRODUCTION

In normal hearing we use both ears, which allows important advantages in interacting with the environment. In spite of this, the auditory information in "high stress" human-machine interface contexts such as aviation is usually received over a monotic (one-ear) headset. It is surprising that, while advanced cab aircraft such as the McDonnell-Douglas MD-88 and the Boeing 767 incorporate highly sophisticated visual displays, cockpit auditory displays are largely a proliferation of semantically unrelated warning sounds which take no advantage of the spatial information which plays so fundamental a role in everyday experience.

Research into improving the human-machine interface is partially motivated by giving attention to operator overload. For example, one source reports that at least 65% of jet transport accidents during 1977-1987 resulted from human errors (Hughes, 1989). One direction for improvement is to access perceptual systems other than vision for communicating important information to an operator. Because spatial hearing is a part of everyday experience that is important for both survival and orientation, it is sensible to determine how it can be manipulated for conveying information in a human-machine interface.

The types of binaural sound manipulation that are feasible to implement depend on the source of the signal. In an aircraft context, there are two distinct types of sources: (1) headphone speech communication using radio transmission originating from ground control or other aircraft, and (2) speech and warning signals that originate from the audio system installed in the cockpit.

For both kinds of signals, binaural sound can improve the intelligibility of speech sources against noise, and assist in the segregation of multiple sound sources. For signals originating in the cockpit, sound spatialization can also be used to organize locations in perceptual auditory space, and to convey urgency or establish redundancy.

This paper reviews both established and evolving techniques for the binaural presentation of sound. Although the example in the application section of the paper makes reference to commercial aircraft cockpits, the results are extendible to air traffic control (ATC), rotorcraft, sonar display, and other multiple-channel human-machine interfaces. The first section reviews of binaural spatialization techniques, along with relevant psychoacoustic considerations for their use. The second section shows how these techniques can be used for improving cockpit auditory displays.

BINAURAL TECHNIQUES AND PSYCHOACOUSTIC CONSIDERATIONS

Spatial hearing refers to the perceived location, size and environmental context of a sound source. In the case of headphone audition, three categories of techniques for manipulating the spatial element of sound are discussed: *lateralization, headphone localization*, and *decorrelation*. These categories are also generally (but not uniquely) distinguished by the kind of audio spatial percept that results from the particular technique.

Lateralization techniques involve manipulation of interaural time and/or intensity differences at each ear; the resulting percept is usually of the sound source moving along the intracranial axis between the ears. Headphone localization techniques simulate spatial hearing in the free field or in an environmental context, essentially by replicating over headphones the interaural differences that occur naturally. In its most successful implementation, the resulting percept can be of an externalized source in three-dimensional space. Decorrelation techniques implement interaural differences using methods other than those mentioned above. In this paper, phase inversion and reverberation techniques are examined.

Lateralization

Lateralization techniques take advantage of two separate mechanisms of the auditory system that are involved in spatial hearing. One mechanism evaluates the amplitude differences at the two ears, while another mechanism evaluates time differences. Human sensitivity to these differences support what is known in the psychoacoustic literature as the "duplex theory" of localization: the differences are abbreviated as interaural level difference (ILD) and interaural time difference (ITD), respectively. It is widely accepted that ILD operates over the entire frequency range, while ITD operates on the fine structure of signals for frequencies below 1.6 kHz, and on the envelope of the signal for frequencies between approximately 200 Hz-20 kHz (Blauert, 1983).

If we present a signal to each speaker of a set of headphones with no ITD or ILD, the sound is heard in the middle of the head. As ITD or ILD is increased past a particular threshold, the sound will begin to shift toward the ear leading in time or greater in amplitude. Once a critical value of ITD or ILD is reached, the sound stops moving along the intracranial axis and remains at the leading or more intense ear. The effective range of ITD is up to approximately 1.5 ms, and the effective range of ILD is around 10 dB. The upper range of ILD is more difficult to determine than ITD: beyond approximately 10 dB, a change in position resulting from ILD is easy to confuse with the corresponding change in auditory extent that occurs around this point (Blauert, 1983). Figure 1 shows these differences rated by subjects on a 1 to 5 scale, where 5 represents maximum displacement. The results shown are valid for speech and noise.

It is relatively easy to implement a ITD/ILD digital signal processing algorithm based on the data in figure 1. For example, consider placing a signal at the extreme left position in the head. We derive the left and right channel outputs $y(n)L$ and $y(n)R$ from an input signal $x(n)$ by multiplying by a gain factor g for ILD:

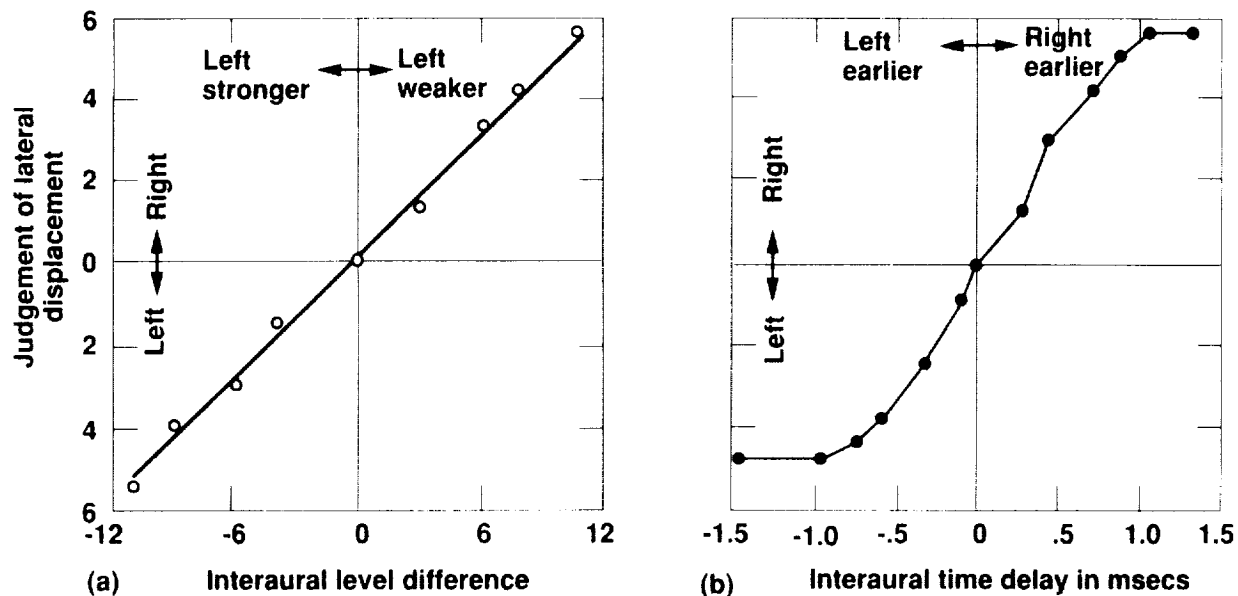$$y(n)L = x(n) \quad y(n)R = x(n) \cdot g, \quad g \cong .3$$

Figure 1. Measurements of interaural differences on lateralization. a) ILD; b) ITD.

and by delaying the input signal by τ for ITD:

$$y(n)L = x(n) \quad y(n)R = x(n + \tau) \quad \tau \cong 1.5 \text{ ms}$$

Figure 2 shows the waveform and circuit diagram displays for these lateralization techniques.

In practical terms, the perceptual effects of ILD and ITD in lateralization can be thought of as additive. Usually, ILD and ITD are pitted dichotically against each other ("traded") in the psychoacoustic literature so as to produce a centered, intracranial position of the perceived sound. Figure 3 illustrates how ILD and ITD can be combined to produce three distinct spatial locations for incoming sounds. In this example, three different radio frequencies are heard in three distinctly different spatial locations: extreme right, extreme left, and center of the head.

The spatial positions attainable using lateralization techniques are severely limited when they are compared to spatial hearing in the real world. Lateralization techniques allow positioning sounds along a line centered between the ears, inside of the head. A more desirable condition would be to have the ability to place sounds outside of the head in any position in three-dimensional space, using headphones.

Headphone Localization

At one time, the term "headphone localization" would be considered an oxymoron. Localization was thought to be only possible with "real world" (nonheadphone) listening, and all headphone listening was assumed to be lateralized. Plenge (1974) was one of the first researchers to demonstrate that localization was indeed possible with headphones. The basis for his argument was based on subjective judgments of externalization of a sound source when recorded with a mannequin head
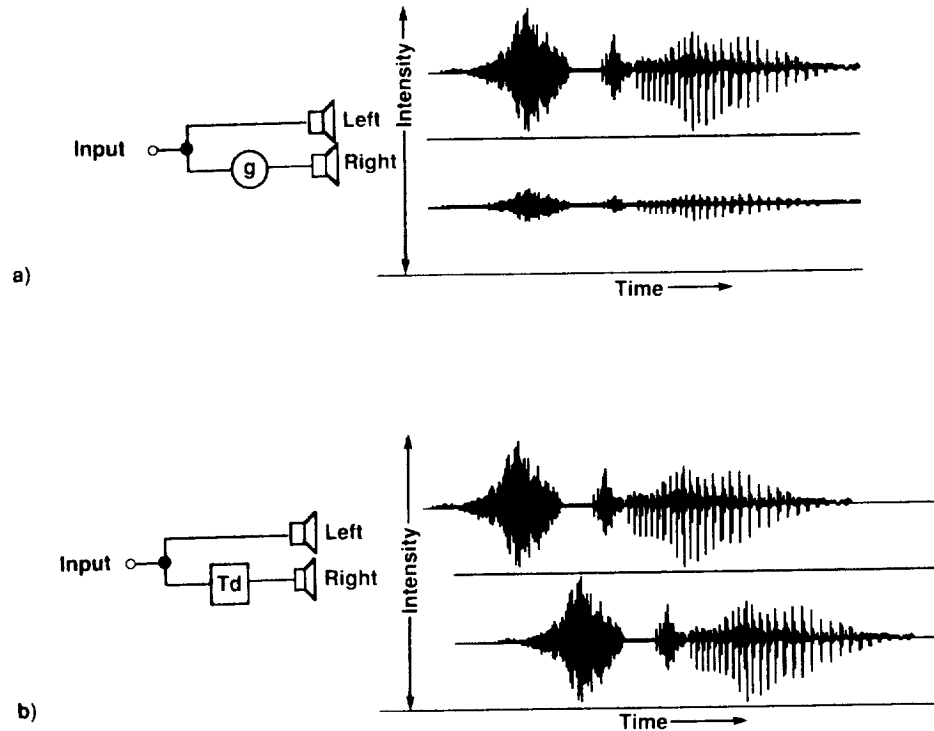
4

Figure 2. Waveform display and circuit diagrams for lateralization techniques. a) ILD; b) ITD.
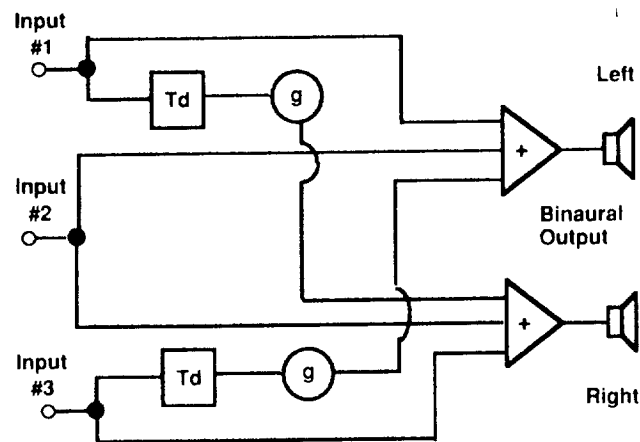
Figure 3. Circuit scheme combining ITD and ILD lateralization techniques to spatially separate three inputs (e.g., radio transmission).

with microphones behind the pinnae. Recent work by Wightman and Kistler (1989b) has reasserted the notion that externalized localization of sounds in three-dimensional space is possible with head-phone listening; their work differs from Plenge in that they substituted digital signal processing techniques for the mannequin head recording process, a technique first used by Platte and Laws (Blauert, 1983). This technique, and its advantages and limitations, are reviewed below.

5

The technique for implementing headphone localization involves creating a digital filter based on measurements of the head related transfer function (HRTF). The HRTF can be thought of as a frequency-dependent amplitude and time delay that results from the resonances of the pinnae and the ear canal, and the effects of head shadowing. These effects combine differentially, as a function of sound source direction; hence, there is a frequency and group delay transfer function imposed on an incoming signal that is unique for any given source position. In other words, the spectrum of the HRTF alters the spectrum of the input signal in a spatially dependent way.

Psychoacoustic research has established the importance of the HRTF for spatial hearing partly because it complements the "duplex theory" of localization. It explains median plane perception of elevation and front-back positions, situations where the ILD and ITD are close to 0 and/or below threshold. Other researchers have shown that localization acuity is diminished overall without pinnae cues (Parker and Oldfield, 1984).

The HRTF is measured by placing a probe microphone close to the eardrum of a subject, or at the entrance to the ear canal (Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989a). The goal is to obtain an impulse response for use in subsequent digital filtering algorithms, and a spectral measurement for purposes of analysis. In simplified terms, an impulse $x(n)$ is sounded from a speaker at a carefully adjusted position in relation to a listener whose head is immobilized. The signal at the microphone $y(n)$ is then recorded, and the procedure is repeated for the desired number of positions. The impulse response of $h(n)$ (the HRTF for that position) is therefore obtained via the convolution of the speaker signal with its path of transmission to the microphone:

$$y(n) = x(n) * h(n)$$

Since $x(n)$ is the the unit-sample impulse ($x(n) = 1$), it follows that $h(n) = y(n)$. By taking the Fourier transform of $h(n)$

$$H\left(e^{j\omega}\right) = \sum_{\eta=-\infty}^{\infty} h(n)e^{-j\omega n}$$

we can obtain the frequency and phase response of the HRTF: in essence

$$H\left(e^{j\omega}\right) = Y\left(e^{j\omega}\right) / X\left(e^{j\omega}\right)$$

Figure 4 shows the magnitude of the HRTF for a single subject for different angles of incidence.

To implement HRTF processing for headphone sound, the spectrum of the incoming sound must be multiplied by the spectra of two HRTF measurements, one for each ear. This is accomplished digitally by the equivalent operation of time-domain convolution, using two finite impulse response (FIR) filters (Oppenheim and Schafer, 1975):

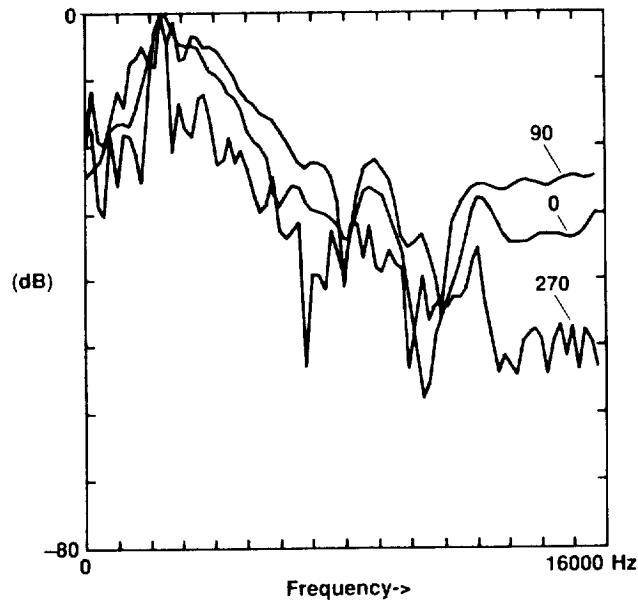$$y(n)L = x(n) * h(n)L; \quad y(n)R = x(n) * h(n)R$$

6

Figure 4. HRTF measurements: one subject, ipsilateral ear, source at 0, 90, and 270° azimuth (based on data from Wightman and Kistler, 1989a).

Figure 5 shows this method. The frequency responses of two filters for synthesizing a source directly opposite the right ear are shown. (Frequency response of the HRTF shown was given by Blauert and then derived by Begault using a digital filter design program (Begault, 1987; Blauert, 1983).)

How successful is the technique, in terms of producing a veridical experience of three-dimensional auditory space? Figure 6 illustrates results obtained by Wightman and Kistler (1989b) that compare free-field and headphone localization performance. The data for both conditions seem in close agreement, but it must be noted that front-back reversals (e.g., mistaking a 0° sound source for a 180° sound source) were corrected in the data analysis. These reversals increased in the headphone case. Based on data from eight subjects, the percentage of front-back confusions from the total number of judgments made with free-field listening is 3-12%, and 6-20% with headphone localization. Individual differences were also marked; the subjects who localized badly over headphones also tended to be the ones that localized badly in free-field conditions.

There is a problem with the technique in that some people are unable to externalize HRTF-processed sound heard through headphones. This is particularly troublesome for simulating sound sources on the median plane; sources synthesized to appear from the front of the listener usually sound as if they are inside of the head. Often, a "bow tie" pattern is perceived when a circle with constant radius from the center of the head is specified (see fig. 7).

We propose three areas for improving headphone localization performance: (1) using a dynamic filter in conjunction with a head-tracking device to allow auditory search and orientation to the virtual sound position, (2) adding environmental cues in the form of reverberation and/or early reflections, and (3) obtaining or synthesizing "optimal" HRTFs.

7

A magnetic head-tracking device that is attached to a set of headphones can transmit three-dimensional coordinates about a listener's head and body position to a receiving device. At Ames, Wenzel and Foster have developed a hardware/software system known as the *Convolvotron* (Wenzel, Wightman, and Foster, 1988) that takes the output coordinates of the tracker and assigns appropriate FIR filters to an input source (see fig. 8). Because the granularity of head movement recorded by the tracker is finer than that of the sampling of the number of positions by the HRTF filters, an interpolation scheme must be used to derive the appropriate impulse response for the FIR filters for intermediate positions.
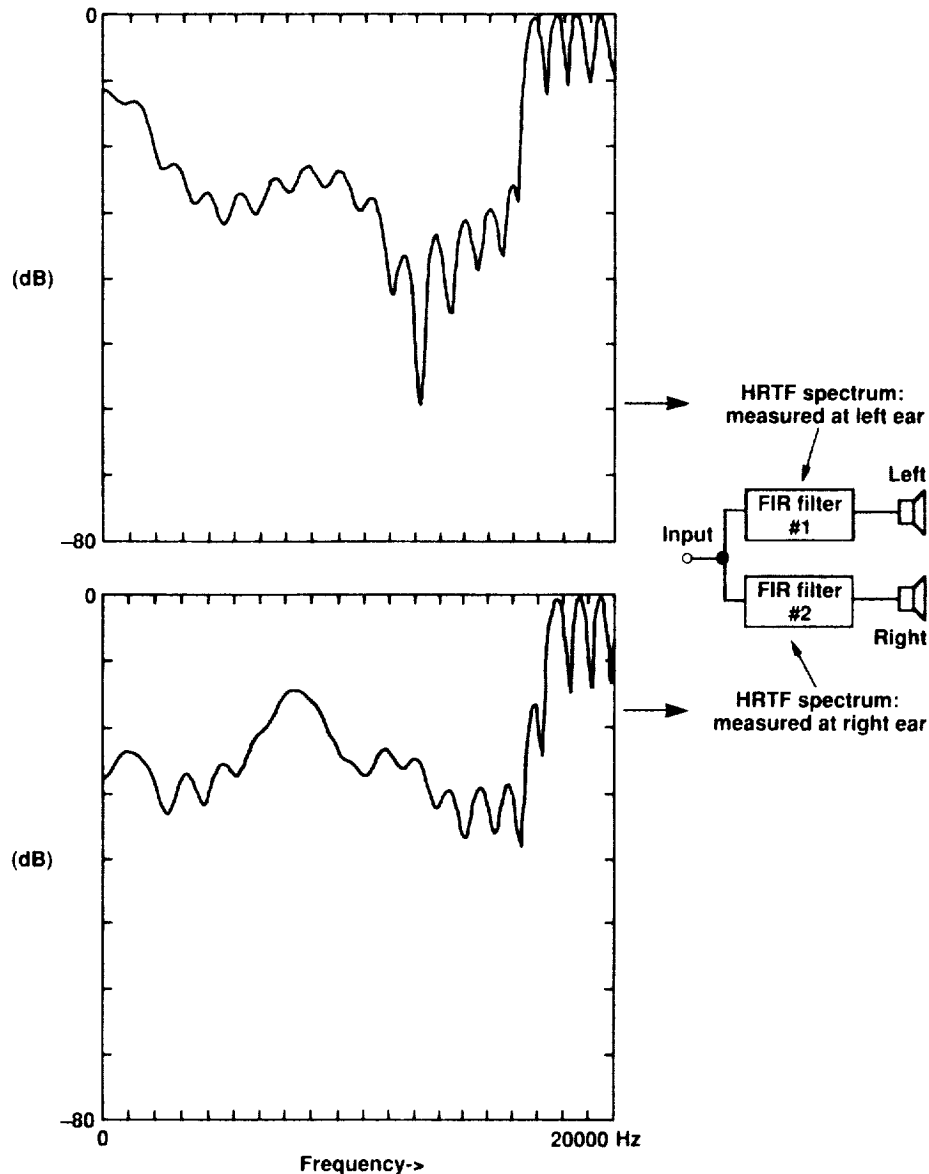
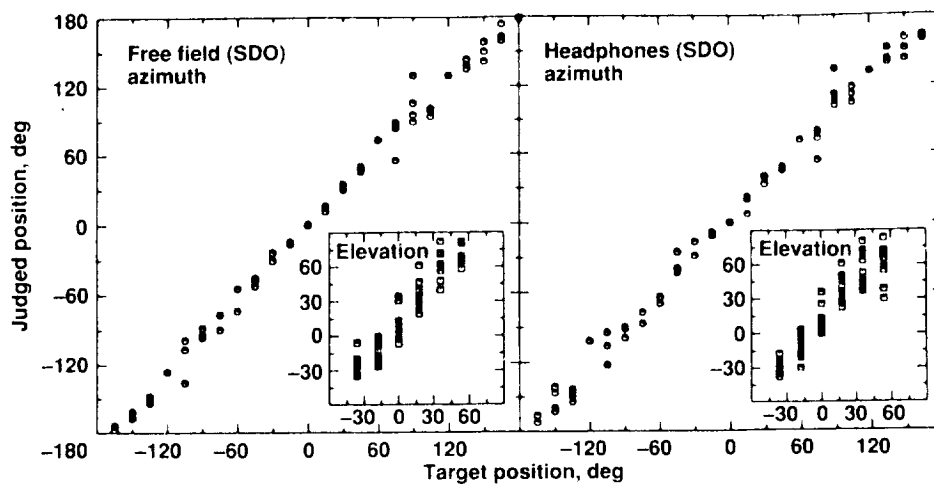Figure 5. Headphone localization technique based on HRTF filtering.

8

Figure 6. Free field and headphone localization performance for a single subject (based on data from Wightman and Kistler, 1989b).
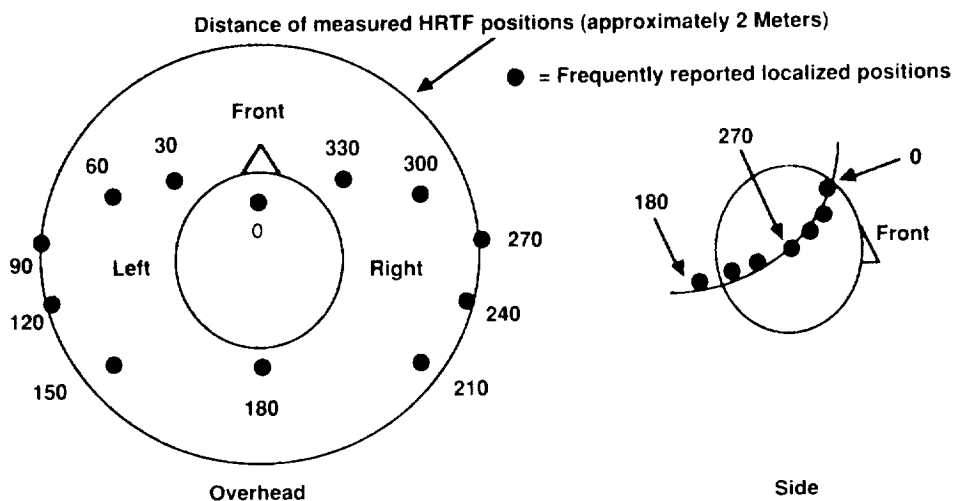

Figure 7. Perceptual result of HRTF filters. Positions frequently reported by subjects when presented with stimuli processed to sound equally distant from the center of the head. Note difficulty in externalizing sounds synthesized to the front and rear.

Another solution to headphone localization errors is to add "environmental" cues. With the method of HRTF filtering described above, the stimulus is convolved with measurements made under anechoic conditions. It lacks environmental information in the form of early reflections and reverberation that also is convolved with a sound source in most listening contexts. Early reflections in normal spatial hearing arrive differentially at the two ears in terms of time of arrival, intensity, and spatial angle of incidence. This situation can be modeled with a ray-tracing computer program that accounts for the dimensions of the enclosure and the position of the listener and sound source. Figure 9 shows one such program Begault developed at Ames. Late reflections (reverberation) may also play a role in externalization and front-back confusions; manipulation of the ratio of direct to reverberant sound has been shown to affect perception of auditory distance (Begault, 1987; von Békésy,
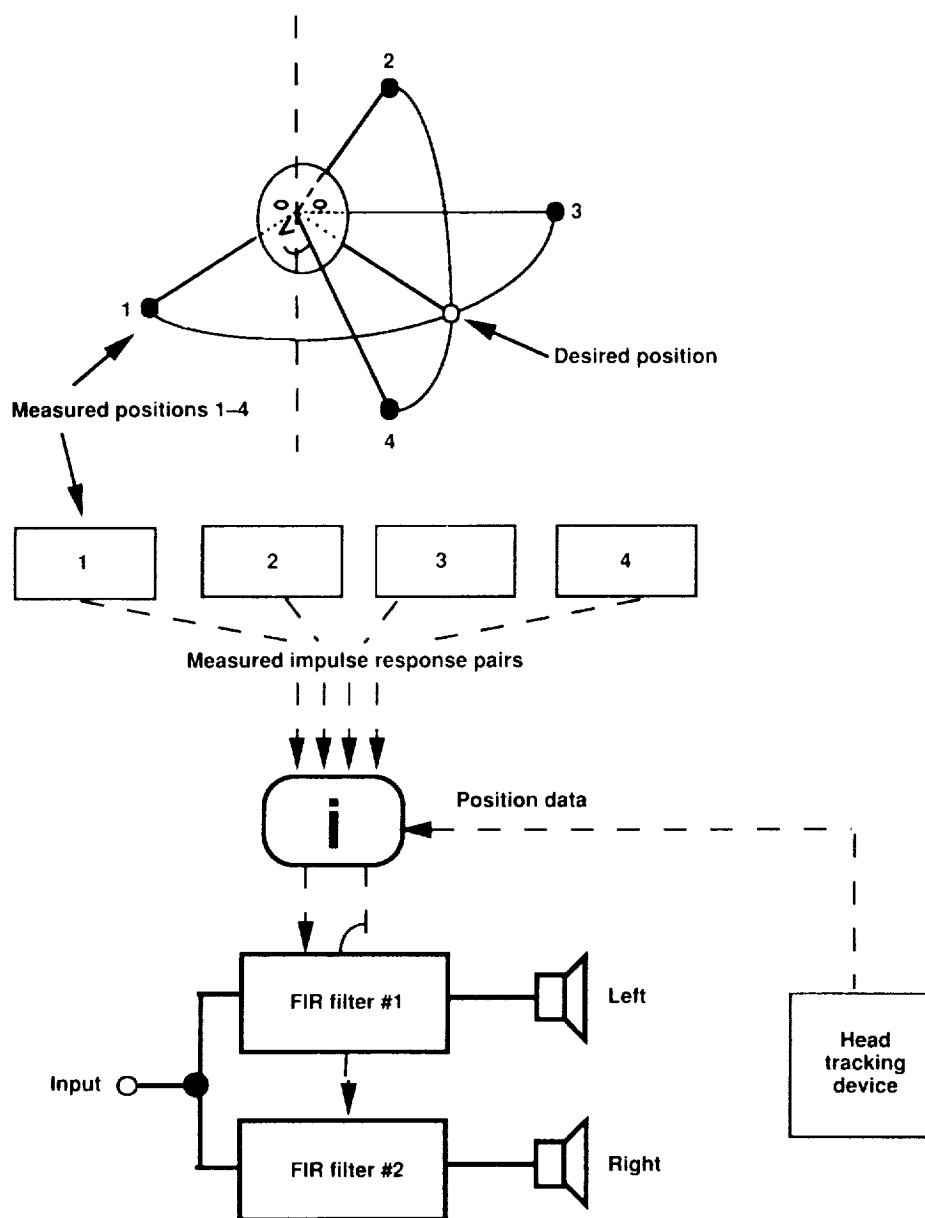
Figure 8. Basic method used by the Convolvotron; circuit shown for one sound source only. HT = head tracker; i = software for interpolating between four pairs of stored HRTF impulse responses.

1960). An increase in the number of synthesized reflections results in a corresponding change in the ratio of energy of the early reflections to the direct sound, and the sound is less spatially correlated in its direction of arrival with respect to the listener.

Another area for improvement has to do with the HRTF itself. If we want to use headphone localization techniques for the general population, can we use a single "set" of HRTF measurements, and if so, how do we derive it? There are essentially two general approaches: averaging methods, and using the HRTFs of a "good localizer." In the first approach, one gathers HRTF data from a

10

Listener: -4.0 -5.0   ori: 270.0
Source: 4.9 4.0   phi: 0.0
Environment file name: xx
   6.0 6.0 (0.5) -6.0 6.0 (0.5) -6.0 6.0 (0.5) 6.0 -6.0 (0.5)
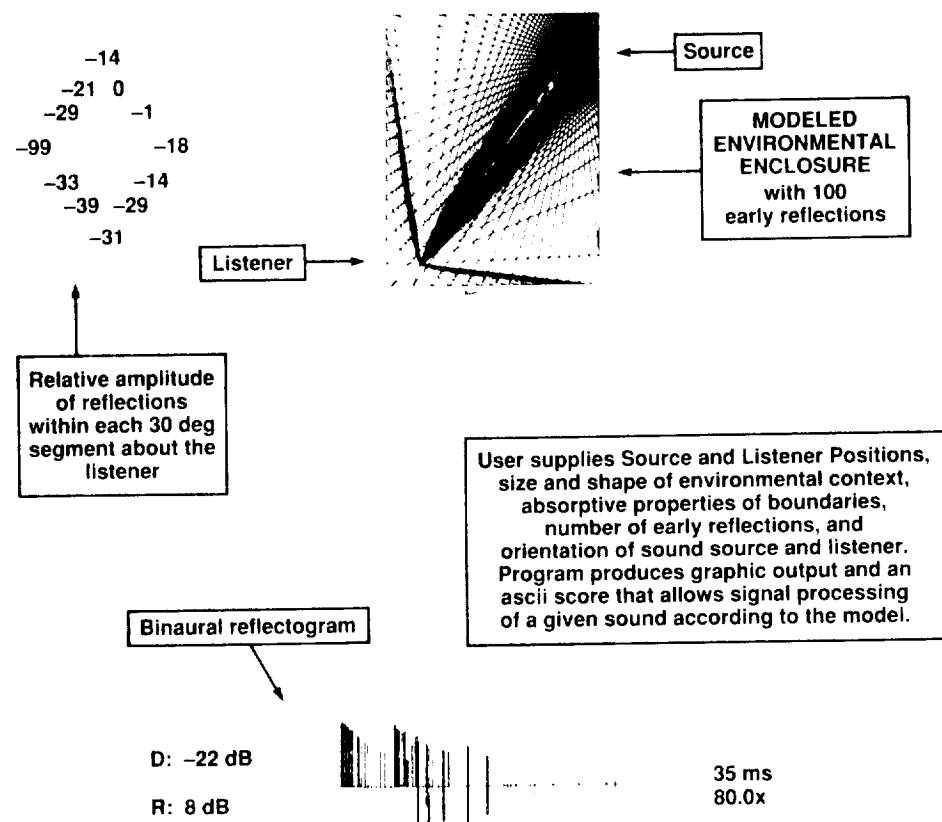No. reflections: 100
Td (i): 0.004 sec.s

Figure 9. Example of user interface for the ray-tracing program used at Ames for synthesizing early reflections.

number of subjects and uses some method of averaging the data. Examples can be found in the work of Blauert and Mehrgardt and Mellert (Blauert, 1983; Mehrgardt and Mellert, 1977). Another form of averaging is to use techniques such as principal components analysis to find significant features within each critical band. The second approach, which we are currently examining at Ames, is to use the HRTFs of a good localizer, i.e., someone whose localization performance both with and without headphones is superior compared to other subjects tested under the same conditions. The problem with the averaging approach is that individual differences in the magnitude and phase characteristics of the transfer function become "smoothed out," resulting in a transfer function with less extreme minima and maxima. The problem with the good localizer approach is not knowing to what degree the subject's performance was idiosyncratic. We anticipate that research using both approaches will help to converge on a set of generalized HRTFs.

11

Decorrelation

Lateralization and headphone localization techniques both transform a single input signal into two different output signals for each ear. This can be viewed as a decorrelation process. However, ILD, ITD, and HRTF filtering are not the only methods for producing binaural decorrelation. Two techniques that can significantly affect both the spatial and timbral dimensions of a sound are phase inversion and multiple delay lines (see fig. 10). While the number of techniques for differentiating two signals from a single one are perhaps innumerable, the use of phase inversion and multiple delay lines is well-known and straightforward in their implementation. Additionally, many decorrelation processes are based on combinations of these techniques.
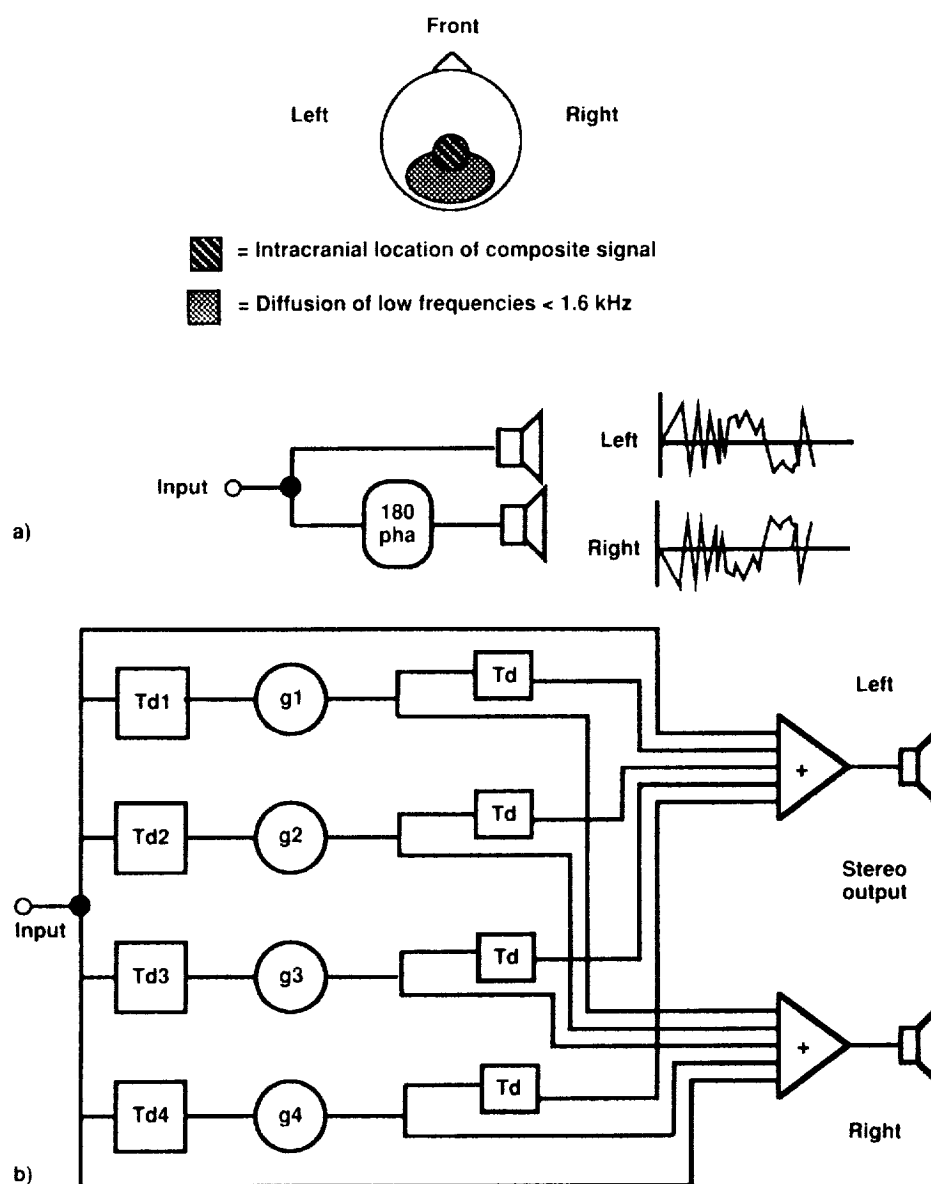

Figure 10. Decorrelation techniques. a) Phase inversion; b) cascaded delay line with interaural shift.

Phase inversion and delay line decorrelation as described here do not affect the percept of the localization of the "center of gravity" of a broadband signal. Rather, it affects the "diffuseness" or spatial extent of a sound source. For example, two sounds can both be localized intracranially, but the decorrelated sound will seem to be more "spread out" than the correlated one (see fig. 10(a)).

With binaural phase inversion of speech, the spreading out effect occurs primarily with the lower frequency components. An explanation for this phenomenon can be given in terms of the frequency selective nature of ITD perception. As mentioned previously, sensitivity to the ITD of the fine structure of a waveform operates only below approximately 1.6 kHz, while sensitivity to the ITD of a waveform envelope operates between 200 Hz-20 kHz (Blauert, 1983). With 180° phase inversion, the envelope of the signals is identical at the two ears, but the fine structure of each harmonic component is time delayed by a half-cycle. Speech contains frequencies within the operating range of both forms of ITD, hence, a differential spreading of sound occurs as a function of frequency.

A multiple delay line is another method of producing two decorrelated signals from a single input. Figure 10(b) illustrates a simplified version of a program available on a typical stereo digital delay device. This circuit adds four time delays to the signal within an approximately 25 ms period, resulting in a timbral change to the signal that is similar to the effect of early reflections heard from walls in a small room. Additionally, by implementing a slight (< 2 ms) time shift between the delay pattern at each channel, the sound image is perceived as being larger or more spread out, as with the phase inversion technique. This process is similar to the interaction of a sound source within an environmental context: timbral changes result because of decorrelated patterns of reflected sound.

APPLICATIONS OF BINAURAL SOUND

Introduction

This section overviews two application areas for binaural sound to aircraft auditory displays. The first is applicable to speech radio transmission, and both apply to speech and warning signals that originate in the cockpit. These areas are *Increased Sensitivity for Communication* (using the binaural advantage over monotic or diotic listening for intelligibility of speech and segregation of a desired source from multiple sources), and *Cognitive Representation of Auditory Space* (using perceived auditory spatial positions as carriers of semantic information). Three types of applications are discussed: urgency (using auditory space to indicate the "levels of urgency" of an auditory warning transmitted with a nonspatial grammar), redundancy (organizing different perceived auditory spatial positions into a grammar that is redundant to a verbal or iconic grammar), and location (the use of perceived location of auditory icons to indicate position of exocentric objects).

Increased Sensitivity for Communication

The difference between a binaural and a diotic system is more profound than simply being an issue of increasing the "pleasing" or "spacious" dimensions of the sound, which is the usual criteria in a commercial application. More important is the fact that using both ears compared to one ear allows a listener important advantages in interacting with the environment, specifically, in suppressing undesirable auditory input such as noise, and to allow localization of sound sources to discrete positions in auditory space.

These two advantages work together to allow perceptual segregation of multiple inputs of sound by the listener. This is exactly what recording engineers do when they mix multitrack recordings: they create a synthetic spatial display of sound sources for two-channel delivery, so that a listener can clearly separate the multiple layers of sound sources. Although two-channel delivery is not the only means of separating sources, it is a powerful one that we use in everyday listening. This binaural advantage can be easily demonstrated by listening with one ear plugged to a person speak in a noisy environment. The noise seems to interfere less with speech when both ears are open. This is in stark contrast to the situation where a pilot must listen to many undifferentiated voices coming over a monotic headset. Indeed, aviation accidents may have occurred because a pilot was unable to attend to the "correct" voice out of the several that he hears coming over the single-channel radio transmission.

Studies by Cherry (1953) and Cherry and Taylor (1954) established the existence of what is commonly called "the cocktail party effect" and its relation to binaural hearing. The term comes from the observation that in a group of people who are all speaking simultaneously, it is still possible to understand a single stream of speech. This has led to a number of studies comparing intelligibility of speech under binaural and monaural presentation. The difference in dB between a masker level (noise or other voices) and the necessary level for intelligibility of the desired signal is measured for both monotic and dichotic conditions. The difference between the two cases, i.e., the improvement in intelligibility due to binaural presentation, is termed the binaural intelligibility level difference, or BILD. Results from experimental data evaluating the BILD differ depending on the stimuli used, and

criteria for evaluating intelligibility; generally, it ranges from 3 to 12 dB (Blauert, 1983). Koening (1950) also described the advantage of binaural over monaural hearing for squelching noise; his results were verified with burst stimuli by Zurek (1979).

Work by Bronkhorst and Plomp (1988) measured the BILD as a function of angle of incidence, using conditions that essentially compare the lateralization techniques discussed earlier (pure ITD or ILD) to headphone localization techniques (listening through the ears of a mannequin head). Figure 11 shows their results; the improvement using headphone localization techniques is around 2-4 dB.
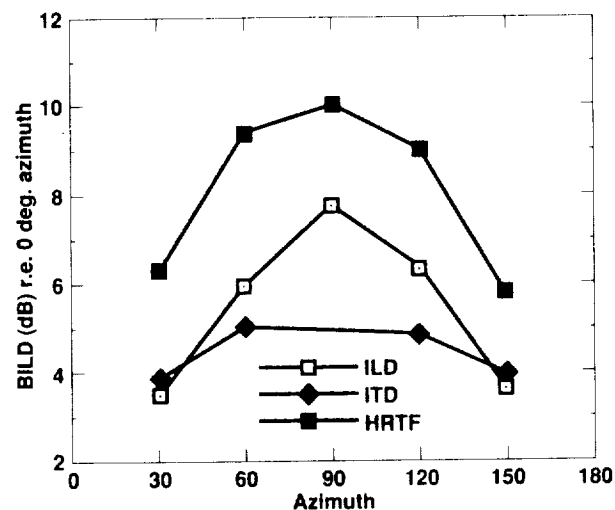

Figure 11. BILD measurements as a function of azimuth: comparison of ILD, ITD, and HRTF-filtered conditions (based on data from Bronkhorst and Plomp, 1988).

Informal studies were conducted at Ames using HRTF filtering techniques. Although the BILD was not measured, the externalization provided by HRTFs with maximal ITD (60, 90, 240 and 300° azimuth) were judged to be superior for selective attention to four voices, compared to ILD or ITD techniques. Additional information can be found in NASA TM-102826.

Because of binaural summation of loudness, a binaurally equipped pilot would also need less amplitude of the signal at each ear, giving an additional advantage in suppressing hearing fatigue. Some pilots object to using headphones on the basis that it isolates them from other crew members. It is possible to alleviate this by using microphones in a monitoring system arrangement. Speech from crew members could be mixed into the overall sound texture of air traffic control (ATC) and cockpit warnings. This has the added advantage that repositioning the body and raising the level of the voice to be heard above background noise (the *Lombard effect*) could be avoided, resulting in less fatigue.

Cognitive Representation of Auditory Space

Introduction– A binaural auditory simulation of sound can potentially define a synthetic space within the "mind's aural eye." We define a "synthetic space" as the listener's organization of the location of sounds within the mind, by using mental imagery. It involves representation that is primarily visual, rather than propositional, in content.

The idea that an analog or imagery-based representational system exists is supported by studies of mental transformations of spatial objects (Cooper and Shepherd, 1978), imagery (Kosslyn, 1981), and cognitive maps (Goldin and Thorndyke, 1982; Kuipers, 1980). Although imagery based on auditory input has not been well studied, experiments concerned with auditory and visual perceptual biases clearly indicate that the two systems interact in the encoding of spatial information (Bertelson and Radeau, 1981).

Spatially correspondent analog representations are also supported by the importance of localization for survival. Spatial pattern recognition would seem to be a fundamental requirement for animal survival in an environment in which many events occur simultaneously in different spatial locations. Especially in the absence of visual stimuli, it is likely that imagery-based modes of thinking occur in connection with spatial hearing. From an applications standpoint, we are interested in eliciting a map of auditory space in the listener so that we can convey semantic meaning as a function of position.

The application of spatial audio to auditory warning systems has been discussed previously in the literature, albeit from many different perspectives. Some experiments manipulated spatial cues in a simplistic manner (e.g., using only interaural intensity differences) or used "impoverished" stimuli, such as pure tones (Mudd, 1965). Doll et al. (1986) concluded from their research that a system of binaural cues for cockpit systems would be beneficial. They used a mannequin head recording technique for auditory spatialization. Examples of applications described below use the lateralization, headphone localization, and decorrelation techniques described previously.

Urgency– A pilot should be able to immediately discern whether a command is urgent well before the command itself is interpreted. An analogy is to the kind of communication parents reserve for their children when they have erred substantially—the children can immediately tell from the pitch and intensity of their parents' voice that they had better pay attention, well before they actually realize the semantic content of the message.

Figure 12 shows a possible spatial audio map for establishing a simple spatial grammar that indicates whether a command is urgent or not. The spatial positions shown could be applied for indicating the relative urgency of a command. The numbered dots represent perceived positions: position 1 was synthesized with a diotic signal, position 2 was synthesized with a dichotic time delay difference, and position 3 was synthesized with a dichotic time delay and amplitude differences.

One possible semantic organization for this set of positions is for the most urgent warning to be processed according to position 1. Less urgent commands are assigned to position 2 or 3. If a sound is transformed from the perceived left side positions to one directly inside of the head, there is a sort of "infringement" of the listener's personal space. It is suggested that, with minimal training, a pilot
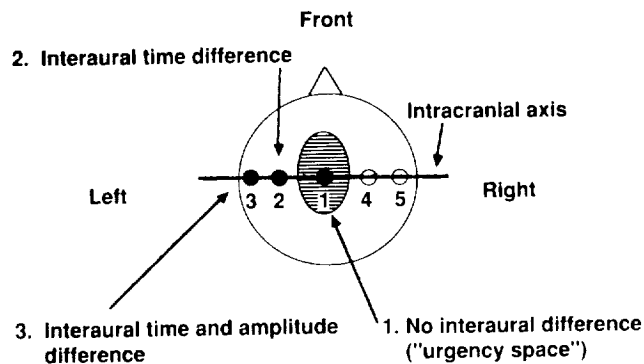
16

Figure 12. A possible topography of perceived auditory spatial positions.

could easily associate a sound inside the head as occupying an *urgency space*, relative to sounds that are perceived to the sides.

Another possibility is that the same spatial topography could be used in relation to repetition of a warning signal. The repetition of a command implies urgency, because the fact that it is repeated means the action has not been taken. For example, the first time a command is given, it comes from position 3 on the left, the second time from position 2 on the left, and the third time from position 1.

By reversing output channels, the signals used to obtain position 2 and 3 can be used to obtain two other positions (4 and 5) at mirror image locations symmetrical from the center of the head. An example of the elaboration of the spatial grammar is as follows: place one class of commands to the left (2 and 3); and one to the right (4 and 5), which reserves position 1 for urgent commands. Using Traffic Collision Avoidance System (TCAS) voice command classifications as an example, traffic commands could come from 2 and 3, nonurgent resolution commands could come from 4 and 5, and urgent resolution commands could come from position 1.

A command originating from the urgency space can be further distinguished by using decorrelation techniques. The multiple delay technique described earlier in section 2 can be applied to sounds from position 1 to underscore their urgency. Urgency as a function of location would in this way be re-emphasized by changes in timbre and image size, further differentiating urgent from nonurgent commands. This use of lateralization and decorrelation techniques to define an urgency space is an application of a *redundancy strategy*.

Auditory icons and redundancy– The cockpit work environment contains a wide range of different auditory input set against a fairly high level of ambient noise from the engine and outside sources. The pilot must extract meaning from vocal communication with other people present in the cockpit, while simultaneously attending to a range of alarms, signals, and automated voice instructions. Depending on the particular on-board system, there are approximately 200-500 different types of warnings that can be potentially displayed to a pilot. In any communication system between source and receiver, redundancy of an intended message assures a better chance of the extraction of a desired signal from noise.

17

There are redundant paradigms where information is encoded more than once. For instance, in the pulse code modulation (PCM) standard for digital encoding of auditory signals, an error correction scheme is used where 30% or more of the digital signal is literally replicated on a tape (Nakajima et al., 1983). Contrasting this is a redundancy paradigm where identical semantics are conveyed with different signs. An everyday example of the latter technique is the use of international traffic signs in the United States (see fig. 13A). Two different types of signs, iconic and verbal, transmit the same message. The icon can be interpreted more quickly than the verbal symbol; the verbal symbol is redundant to reduce the possibility of misinterpretation.

A similar method is possible for including redundancy in TCAS messages. The technique is to link each verbal message with an appropriate auditory icon, and then precede the verbal message with the icon upon presentation (see fig. 13B). The advantage to this is that the semantic content can be grasped quickly by recognizing the auditory icon, and can then be verified by the verbal icon. This idea has been suggested in several studies; for example, Patterson (1982) suggests a technique where verbal commands are "sandwiched" temporally between auditory icons.
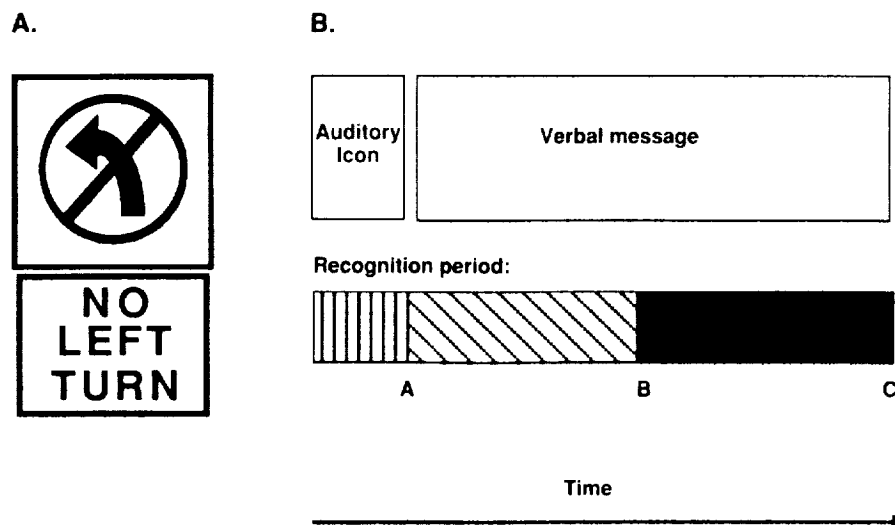

Figure 13. Identical semantic content achieved with combinations of nonverbal and verbal semiology.

Redundancy has been discussed so far in the context of a sequential, differentiated process of communication, where two types of semiotic objects follow one another in time. But redundancy can also be established by presenting a pilot with multiple percepts that parallel one another during the same time interval. Multi sensory cues have begun to be used as redundant indicators in some advanced aircraft. For example, some modern aircraft use visual, auditory, and tactile cues (red light, warning horn, and a stick shaker) for indicating aircraft engine stall. The supposition is that by accessing multiple perceptual pathways to the pilot at the same time, the chance of error is reduced (Mowbray and Gebhard, 1961).

It is also possible to achieve parallel redundancy in a purely aural manner. The manipulation of auditory space for this purpose is possible as long as the semiology of position is somewhat limited. Returning to the concept of indicating urgency by spatialization to a center position (position 1 in

figure 12), the "invasion" of a sound into urgency space can be combined with an appropriately urgent auditory icon. In this case both the icon and the spatial position of the sound simultaneously warn a pilot of urgency. Hence, a limited grammar of spatial sound manipulation could be used with urgent aural icons to further establish an urgency space.

Location of auditory cues in relation to exocentric objects– Perhaps the most advanced use of spatial cues is to implement a headphone localization system where the location of an object in three-dimensional space is paralleled by its perceived auditory position. Related research is under way at Ames' virtual environment project VIEW (Fisher et al., 1988). In a telerobotics application of VIEW, an example would be to use spatialized auditory cues to indicate the position of objects near a robot, such as a spacecraft. With three-dimensional sound, the position of objects associated with these auditory cues can be monitored without having to capture their position visually.

Headphone localization is applicable to cockpit warning systems that require communication of positional information. Even a sound display limited to one elevation (e.g., at head level) with a sampling of positions along the 360° azimuth would be useful in a cockpit environment. For example, a pilot maneuvering through a busy airport approach would be able to use an auditory spatial grammar based on the 12 hour hand clock positions, conceiving themselves in the center of the clock with 12 at 0° azimuth and 6 at 180° azimuth. If a pilot heard a TCAS warning that stated, "traffic at 8 o'clock," a headphone localization technique could provide redundancy by placing the perceived position at 120° azimuth. The problems with front-back reversals mentioned previously limits the ability. It is still currently unfeasible to relate all 12 clock positions in a virtual space to perceived positions with absolute confidence.

With continued research, the externalization and front-back reversal problems inherent in headphone localization should be alleviated. It is definitely possible to use a limited number of HRTF-synthesized positions, especially for conveying positions to the sides; their effect has been found by several authors as an improvement over traditional intensity or time difference stereo techniques for creating spatial auditory displays (Begault, 1986; Griesinger, 1989; Kendall and Martens, 1984).

CONCLUDING REMARKS

Three techniques for controlling the binaural sound image—lateralization, headphone localization, and decorrelation—have been discussed in terms of implementation, perception, and current areas of research, including efforts at Ames. This was exemplified by showing an application of binaural sound to cockpit speech and warning systems. Improvements in the intelligibility of speech and segregation of multiple source inputs were outlined. In addition, suggestions were made for the use of the mapping of auditory space to convey urgency, redundancy, and position within an auditory display. Using the techniques outlined would provide powerful control over a vivid perceptual faculty.

REFERENCES

Begault, D. R.: Spatial Manipulation and Computers. Ex Tempore, vol. IV, no. 1, 1986, pp. 56-88.

Begault, D. R.: Control of Auditory Distance. Dissertation, University of California San Diego, La Jolla, CA, 1987.

Bertelson, P.; and Radeau, M.: Cross-Modal Bias and Perceptual Fusion with Auditory-Visual Spatial Discordance. Perception and Psychophysics, vol. 29, 1981, pp. 578-584.

Blauert, J.: Spatial Hearing: the Psychophysics of Human Sound Localization. Cambridge:MIT Press, 1983.

Bronkhorst, A. W.; and Plomp, R.: The Effect of Head-Induced Interaural Time and Level Differences on Speech Intelligibility in Noise. J. Acous. Soc. America, vol. 83, no. 4, 1988, pp. 1508-1516.

Cherry, E. C.: Some Experiments on the Recognition of Speech with One and with Two Ears. J. Acous. Soc. America, vol. 25, no. 5, 1953, pp. 975-979.

Cherry, E. C.; and Taylor, W. K.: Some Further Experiments on the Recognition of Speech with One and Two Ears. J. Acous. Soc. America, vol. 26, no. 4, 1954, pp. 549-554.

Cooper, L. A.; and Shephard, R. N.: Transformations on Representations of Objects in Space. In E. C. Carterrette and M. P. Friedman (Eds.), Handbook of Perception. New York:Academic Press, 1978.

Doll, T. J.; Gerth, J. M.; Engelman, W. R.; and Folds, D. J.: Development of Simulated Directional Audio for Cockpit Applications. Tech. Report AAMRL-TR-86-014, Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory, 1986.

Fisher, S. S.; Wenzel, E. M.; Coler, C.; and McGreevy, M. W.: Virtual Interface Environment Workstations. In Proceedings of the Human Factors Society 32nd Annual Meeting, Anaheim, CA, 1988, pp. 91-95.

Goldin, S. E.; and Thorndyke, P. W.: Simulating Navigation for Spatial Knowledge Acquisition. Human Factors, vol. 24, 1982, pp. 457-471.

Griesinger, D.: Equalization and Spatial Equalization of Dummy Head Recordings for Loudspeaker Reproduction. J. Audio Eng. Soc., vol. 37, nos. 1-2, 1989, pp. 20-29.

Hughes, D.: Glass Cockpit Study Reveals Human Factors Problems. Aviat. Week and Space Tech., vol. 131, no. 6, 1989, pp. 32-36.

Kendall, G. S.; and Martens, W. L.: Simulating the Cues of Spatial Hearing in Natural Environments. Tech. Report, Northwestern University, 1984.

Koening, W.: Subjective Effects in Binaural Hearing. J. Acous. Soc. America, vol. 22, no. 1, 1950, pp. 61-62.

Kosslyn, S. M.: The Medium and the Message in Mental Imagery: A Theory. Psychological Rev., vol. 88, 1981, pp. 46-66.

Kuipers, B. J.: The Cognitive Map: Could it Have Been Any Other Way? In H. L. Pick Jr. and L. P. Acredolo (Eds.), Spatial Orientation: Theory, Research, and Applications. Plenum Press, 1980.

Mehrgardt, S.; and Mellert, V.: Transformation Characteristics of the External Human Ear. J. Acous. Soc. America, vol. 61, no. 6, 1977, pp. 1567-1576.

Mowbray, G. H.; and Gebhard, J. W.: Man's Senses as Information Channels. In H. W. Sinaiko (Ed.), Selected Papers on Human Factors in the Design and Use of Control Systems. New York: Dover, 1961.

Mudd, S. A.: Experimental Evaluation of Binary Pure-Tone Audition. J. Appl. Psych., vol. 49, no. 2, 1965, pp. 112-121.

Nakajima, D.; Doi, T.; Fukuda, J.; and Iga, A.: Digital audio technology. Blue Ridge Summit, PA: Tab Books, 1983.

Oppenheim, A. V.; and Schafer, R. W.: Digital signal processing. Englewood Cliffs, NJ: Prentice-Hall, 1975.

Parker, S. P. A.; and Oldfield, S. R.: Acuity of Sound Localization: A Topography of Sound Space. II. Pinnae Cues Absent. Perception, vol. 13, 1984, pp. 600-617.

Patterson, R.: Guidelines for Auditory Warning Systems on Civil Aircraft. Tech. Rep. 82017. Civil Aviation Authority, London, 1982.

Plenge, G.: On the Differences Between Localization and Lateralization. J.Acous. Soc. America, vol. 56, no. 3, 1974, pp. 944-951.

von Békésy, Georg: Experiments in Hearing. New York:McGraw-Hill, 1960.

Wenzel, E. M.; Wightman, F. L.; and Foster, S. H.: A Virtual Display System for Conveying Three-Dimensional Acoustic Information. In Proceedings of the Human Factors Society 32nd Annual Meeting. Anaheim, CA, 1988.

Wightman, F. L.; and Kistler, D. J.: Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis. J. Acous. Soc. America, vol. 85, no. 2, 1989a, pp. 858-867.

Wightman, F. L.; and Kistler, D. J.: Headphone Simulation of Free-Field Listening. II: Psychophysical Validation. J. Acous. Soc. America, vol. 85, no. 2, 1989b, pp. 868-878.

Zurek, P. M.: Measurements of Binaural Echo Suppression. J. Acous. Soc. America, vol. 66, no. 6, 1979, pp. 1750-1757.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No. NASA TM-102279	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces		5. Report Date August 1990
		6. Performing Organization Code
7. Author(s) Durand R. Begault and Elizabeth M. Wenzel		8. Performing Organization Report No. A-90066
		10. Work Unit No. 505-69-01
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035		11. Contract or Grant No.
		13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001		14. Sponsoring Agency Code

15. Supplementary Notes

Point of Contact: Elizabeth M. Wenzel, Ames Research Center, MS 239-3,
Moffett Field, CA 94035-1000
(415)604-6290 or FTS 464-6290

16. Abstract

The implementation of binaural sound to speech and auditory sound cues ("auditory icons") is addressed from both an applications and technical standpoint. Techniques overviewed include processing by means of filtering with head-related transfer functions. Application to advanced cockpit human interface systems is discussed, although the techniques are extendable to any human-machine interface. Research issues pertaining to three-dimensional sound displays under investigation at the Aerospace Human Factors Division at NASA Ames Research Center are described.

17. Key Words (Suggested by Author(s)) Psychoacoustics Digital signal processing Human-machine interfaces	18. Distribution Statement Unclassified–Unlimited Subject Category–53		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 28	22. Price A03