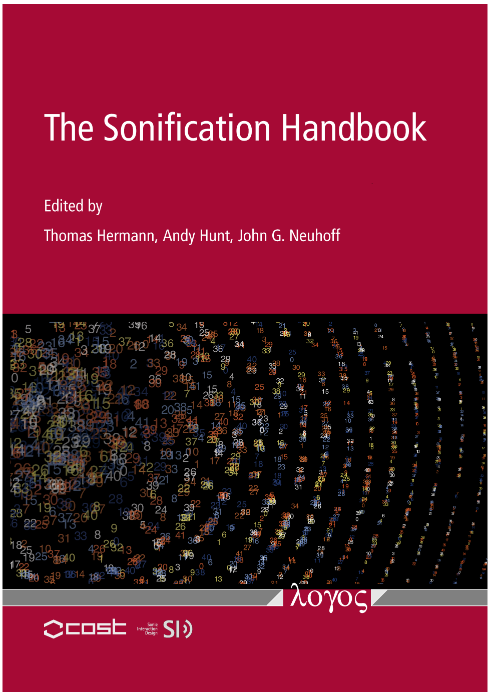

The Sonification Handbook

Edited by
Thomas Hermann, Andy Hunt, John G. Neuhoff

Logos Publishing House, Berlin, Germany
ISBN 978-3-8325-2819-5
2011, 586 pages

Reference:
Hermann, T., Hunt, A., Neuhoff, J. G., editors (2011). *The Sonification Handbook*. Logos Publishing House, Berlin, Germany.

Chapter 4

Perception, Cognition and Action in Auditory Display

John G. Neuhoff

This chapter covers auditory perception, cognition, and action in the context of auditory display and sonification. Perceptual dimensions such as pitch and loudness can have complex interactions, and cognitive processes such as memory and expectation can influence user interactions with auditory displays. These topics, as well as auditory imagery, embodied cognition, and the effects of musical expertise will be reviewed.

Reference:
Neuhoff, J. G. (2011). Perception, cognition and action in auditory display. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, The Sonification Handbook, chapter 4, pages 63–85. Logos Publishing House, Berlin, Germany.

Chapter 4

Perception, Cognition and Action in Auditory Displays

John G. Neuhoff

4.1 Introduction

Perception is almost always an automatic and effortless process. Light and sound in the environment seem to be almost magically transformed into a complex array of neural impulses that are interpreted by the brain as the subjective experience of the auditory and visual scenes that surround us. This transformation of physical energy into "meaning" is completed within a fraction of a second. However, the ease and speed with which the perceptual system accomplishes this Herculean task greatly masks the complexity of the underlying processes and often times leads us to greatly underestimate the importance of considering the study of perception and cognition, particularly in applied environments such as auditory display.

The role of perception in sonification has historically been of some debate. In 1997 when the International Community for Auditory Display (ICAD) held a workshop on sonification, sponsored by the National Science Foundation, that resulted in a report entitled "*Sonification Report: Status of the Field and Research Agenda*" (Kramer, et al., 1999). One of the most important tasks of this working group was to develop a working definition of the word "sonification". The underestimation of the importance of perception was underscored by the good deal of discussion and initial disagreement over including anything having to *do* with "perception" in the definition of sonification. However, after some debate the group finally arrived at the following definition:

> "*...sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation.*"

The inclusion of the terms "*perceived relations*" and "*communication or interpretation*" in this definition highlights the importance of perceptual and cognitive processes in the development of effective auditory displays. Although the act of perceiving is often an effortless and automatic process it is by no means simple or trivial. If the goal of auditory display is to convey meaning with sound, then knowledge of the perceptual processes that turn sound into meaning is crucial.

No less important are the cognitive factors involved in extracting meaning from an auditory display and the actions of the user and interactions that the user has with the display interface. There is ample research that shows that interaction, or intended interaction with a stimulus (such as an auditory display) can influence perception and cognition.

Clearly then, an understanding of the perceptual abilities, cognitive processes, and behaviors of the user are critical in designing effective auditory displays. The remainder of this chapter will selectively introduce some of what is currently known about auditory perception, cognition, and action and will describe how these processes are germane to auditory display.

Thus, the chapter begins with an examination of "low level" auditory dimensions such as pitch, loudness and timbre and how they can best be leveraged in creating effective auditory displays. It then moves to a discussion of the perception of auditory space and time. It concludes with an overview of more complex issues in auditory scene analysis, auditory cognition, and perception action relationships and how these phenomena can be used (and misused) in auditory display.

4.2 Perceiving Auditory Dimensions

There are many ways to describe a sound. One might describe the sound of an oboe by its timbre, the rate of note production, or by its location in space. All of these characteristics can be referred to as "auditory dimensions". An auditory dimension is typically defined as the subjective perceptual experience of a particular physical characteristic of an auditory stimulus. So, for example, a primary physical characteristic of a tone is its fundamental frequency (usually measured in cycles per second or Hz). The perceptual dimension that corresponds principally to the physical dimension of frequency is "pitch", or the apparent "highness" or "lowness" of a tone. Likewise the physical intensity of a sound (or its amplitude) is the primary determinant of the auditory dimension "loudness".

A common technique for designers of auditory displays is to use these various dimensions as "channels" for the presentation of multidimensional data. So, for example, in a sonification of real-time financial data Janata and Childs (2004) used rising and falling pitch to represent the change in price of a stock and loudness to indicate when the stock price was approaching a pre-determined target (such as its thirty day average price). However, as is made clear in the previous chapter on psychoacoustics, this task is much more complex than it first appears because there is not a one-to-one correspondence between the physical characteristics of a stimulus and its perceptual correlates. Moreover, (as will be shown in subsequent sections) the auditory dimensions "interact" such that the pitch of a stimulus can influence its loudness, loudness can influence pitch, and other dimensions such as timbre and duration can all influence each other. This point becomes particularly important in auditory display, where various auditory dimensions are often used to represent different variables in a data set. The complexities of these auditory interactions have yet to be fully addressed by the research community. Their effects in applied tasks such as those encountered in auditory display are even less well illuminated. However, before discussing how the various auditory dimensions interact, the discussion turns toward three of the auditory dimensions that are most commonly used in auditory display: pitch, loudness, and timbre.

4.2.1 Pitch

Pitch is perhaps the auditory dimension most frequently used to represent data and present information in auditory displays. In fact, it is rare that one hears an auditory display that does not employ changes in pitch. Some of the advantages of using pitch are that it is easily manipulated and mapped to changes in data. The human auditory system is capable of detecting changes in pitch of less than 1Hz at a frequency of 100Hz (See Chapter 3 section 3.4 of this volume). Moreover, with larger changes in pitch, musical scales can provide a pre-existing cognitive structure that can be leveraged in presenting information. This would occur for example in cases where an auditory display uses discrete notes in a musical scale to represent different data values.

However, there are a few disadvantages in using pitch. Some work suggests that there may be individual differences in musical ability that can affect how a display that uses pitch change is perceived (Neuhoff, Kramer, & Wayand, 2002). Even early psychophysicists acknowledged that musical context can affect pitch perception. The revered psychophysicist S.S. Stevens, for example, viewed the intrusion of musical context into the psychophysical study of pitch as an extraneous variable. He tried to use subjects that were musically naive and implemented control conditions designed to prevent subjects from establishing a musical context. For example instead of using frequency intervals that corresponded to those that followed a musical scale (e.g., the notes on a piano), he used intervals that avoided any correspondence with musical scales. In commenting about the difficulty of the method involved in developing the mel scale (a perceptual scale in which pitches are judged to be equal in distance from one another), Stevens remarked "*The judgment is apparently easier than one might suppose, especially if one does not become confused by the recognition of musical intervals when he sets the variable tone.*" (Stevens & Davis, 1938, p. 81). It was apparent even to Stevens and his colleagues then that there are privileged relationships between musical intervals that influence pitch perception. In other words, frequency intervals that correspond to those that are used in music are more salient and have greater "meaning' than those that do not, particularly for listeners with any degree of musical training.

If pitch change is to be used by a display designer, the changes in pitch must be mapped in some logical way to particular changes in the data. The question of mapping the direction of pitch change used in a display (rising or falling) to increasing or decreasing data value is one of "polarity". Intuitively, increases in the value of a data dimension might seem as though they should be represented by increases in the pitch of the acoustic signal. Indeed many sonification examples have taken this approach. For example, in the sonification of historical weather data, daily temperature has been mapped to pitch using this "positive polarity", where high frequencies represent high temperatures and low frequencies represent low temperatures (Flowers, Whitwer, Grafel, & Kotan, 2001). However, the relationship between changes in the data value and frequency is not universal and in some respects depends on the data dimension being represented and the nature of the user. For example, a "negative polarity" works best when sonifying size, whereby decreasing size is best represented by *increasing* frequency (Walker, 2002). The cognitive mechanisms that underly polarity relationships between data and sound have yet to be investigated.

Walker and colleagues (Walker 2002; Walker 2007; Smith & Walker, 2002; Walker & Kramer, 2004) have done considerable work exploring the most appropriate polarity and conceptual mappings between data and sound dimensions. This work demonstrates the complexity of the problem of mapping pitch to data dimensions with respect to polarity. Not only do different data dimensions (e.g., temperature, size, and pressure) have different effective polarities, but there are also considerable individual differences in the choice of preferred polarities. Some users even show very little consistency in applying a preferred polarity (Walker, 2002). In other cases distinct individual differences predict preferred polarities. For example, users with visual impairment sometimes choose a polarity that is different from those without visual impairment (Walker & Lane, 2001).

In any case, what may seem like a fairly simple auditory dimension to use in a display has some perhaps unanticipated complexity. The influence of musical context can vary from user to user. Polarity and scaling can vary across the data dimensions being represented. Mapping data to pitch change should be done carefully with these considerations in the forefront of the design process.

4.2.2 Loudness

Loudness is a perceptual dimension that is correlated with the amplitude of an acoustic signal. Along with pitch, it is easily one of the auditory dimensions most studied by psychologists and psychoacousticians. The use of loudness change in auditory displays, although perhaps not as common as the use of pitch change, is nonetheless ubiquitous. The primary advantages of using loudness change in an auditory display are that it is quite easy to manipulate, and is readily understood by most users of auditory displays. However, despite its frequent use, loudness is generally considered a poor auditory dimension for purposes of representing continuous data sets. There are several important drawbacks to using loudness change to represent changes in data in sonification and auditory display.

- First, the ability to discriminate sounds of different intensities, while clearly present, lacks the resolution that is apparent in the ability to discriminate sounds of different frequencies.
- Second, memory for loudness is extremely poor, especially when compared to memory for pitch.
- Third, background noise and the sound reproduction equipment employed in any given auditory display will generally vary considerably depending on the user's environment. Thus, reliable sonification of continuous variables using loudness change becomes difficult (Flowers, 2005).
- Finally, there are no pre-existing cognitive structures for loudness that can be leveraged in the way that musical scales can be utilized when using pitch. Loudness, like most other perceptual dimensions, is also subject to interacting with other perceptual dimensions such as pitch and timbre.

Nonetheless, loudness change is often used in auditory display and if used correctly in the appropriate contexts, it can be effective. The most effective use of loudness change usually occurs when changes in loudness are constrained to two or three discrete levels that are mapped to two or three discrete states of the data being sonified. In this manner, discrete changes in loudness can be used to identify categorical changes in the state of a variable or to indicate when a variable has reached some criterion value. Continuous changes in loudness can be used to sonify trends in data. However, the efficacy of this technique leaves much to be desired. Absolute data values are particularly difficult to perceive by listening to loudness change alone. On the other hand, continuous loudness change can be mapped *redundantly* with changes in pitch to enhance the salience of particularly important data changes or auditory warnings. This point will be expanded below when discussing the advantageous effects of dimensional interaction.

4.2.3 Timbre

Timbre *(pronounced TAM-bur)* is easily the perceptual dimension about which we have the least psychophysical knowledge. Even *defining* timbre has been quite a challenge. The most often cited definition of timbre (that of the American National Standards Institute or ANSI) simply identifies what timbre is *not* and that whatever is left after excluding these characteristics– is timbre. ANSI's "negative definition" of timbre reads like this: "*...that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different*". In other words, timbre is what allows us to tell the difference between a trumpet and a clarinet when both are playing the same pitch at the same loudness. Part of the difficulty in defining timbre stems from the lack of a clear physical stimulus characteristic that is ultimately responsible for the perception of timbre. Unlike the physical-perceptual relationships of amplitude-loudness and frequency-pitch, there is no single dominant physical characteristic that correlates well with timbre. The spectral profile of the sound is most often identified as creating the percept of timbre, and spectrum does indeed influence timbre. However, the time varying characteristics of the amplitude envelope (or attack, sustain and decay time of the sound) has also been shown to have a significant influence on the perception of timbre.

Timbre can be an effective auditory dimension for sonification and has been used both as a continuous and a categorical dimension. Continuous changes in timbre have been proposed for example, in the auditory guidance of surgical instruments during brain surgery (Wegner, 1998). In this example, a change in spectrum is used to represent changes in the surface function over which a surgical instrument is passed. A homogeneous spectrum is used when the instrument passes over a homogeneous surface, and the homogeneity of the spectrum changes abruptly with similar changes in the surface area. Alternatively, discrete timbre changes, in the form of different musical instrument sounds can be used effectively to represent different variables or states of data. For example, discrete timbre differences have been used to represent the degree of confirmed gene knowledge in a sonification of human chromosome 21 (Won, 2005). Gene sequence maps are typically made in six colors that represent the degree of confirmed knowledge about the genetic data. Won (2005) employed six different musical instruments to represent the various levels of knowledge. When using different timbres it is critical to choose timbres that are easily discriminable. Sonification using similar timbres can lead to confusion due to undesirable perceptual grouping (Flowers, 2005).

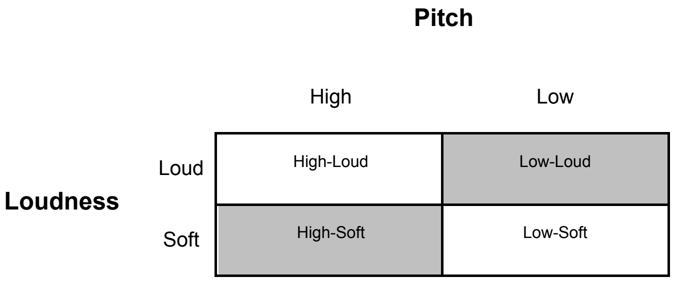

Figure 4.1: Schematic diagram of the four types of stimuli used in a speeded sorting task to test dimensional interaction. Grey and white boxes indicate “incongruent” and “congruent” stimuli respectively.

4.2.4 Interacting Perceptual Dimensions

At first glance, it would be easy to believe that distinct changes in individual acoustic characteristics of a stimulus such as frequency, intensity and spectrum would be perceived as perceptually distinct characteristics of a sound. However, there is growing evidence to the contrary. Changes in acoustic dimensions affect not only the percept of the corresponding perceptual dimension, but also specific perceptual characteristics of *other* perceptual dimensions. In other words, changes in one dimension (such as pitch) can affect perceived changes in the others (such as loudness). Given that the auditory system has evolved in an environment where stimuli constantly undergo simultaneous dynamic change of multiple acoustic parameters, perhaps this should come as no surprise. However, the implications of this kind of dimensional interaction for sonification and auditory display are important.

Perception researchers have devised a set of “converging operations” that are used to examine interacting perceptual dimensions (Garner, 1974). Listeners are typically presented with stimuli that vary along two dimensions such as pitch and loudness. They are instructed to attend to one dimension (e.g., pitch) and ignore changes in the other. In a speeded sorting task, for example, listeners would be presented with four types of sounds, with pitch and loudness each having two values (See Figure 4.1). The 2 (pitch) × 2 (loudness) matrix yields four sounds that are 1.) “high-loud”, 2.) “high-soft”, 3.) “low-loud”, and 4.) “low soft”. Listeners might be asked to perform a two-alternative forced-choice task in which they are asked to ignore loudness and simply press one of two buttons to indicate whether the pitch of the sound is “high” or “low”. The researcher measures the amount of time required to make each response and the number of errors in each condition. Results typically show that responses are faster and more accurate in the “congruent” conditions of “high-loud” and “low-soft” than in the incongruent conditions. Because performance in the attended dimension is affected by variation in the *unattended* dimension, the two dimensions are said to interact. Pitch, timbre, loudness, and a number of other perceptual dimensions commonly used by display designers have all been shown to interact perceptually.

Thus, simply mapping orthogonal variables to different parts of the acoustic signal does not guarantee that they will remain orthogonal perceptually (Anderson & Sanderson, 2009;

Melara & Marks, 1990; Neuhoff, Kramer & Wayand, 2002, Walker & Ehrenstein, 2000), and therein lies a potential problem for designers of auditory displays. On the other hand, the "problem" of interacting perceptual dimensions has also been capitalized upon by redundantly mapping multiple perceptual dimensions (e.g., pitch and loudness) to a single data variable. This technique makes changes in the data more salient and is particularly effective for important signals. Depending on the context of the display, dimensional interaction can have both detrimental and advantageous effects in the context of auditory display. Each of these effects will now be explored.

4.2.5 Detrimental Effects of Dimensional Interaction

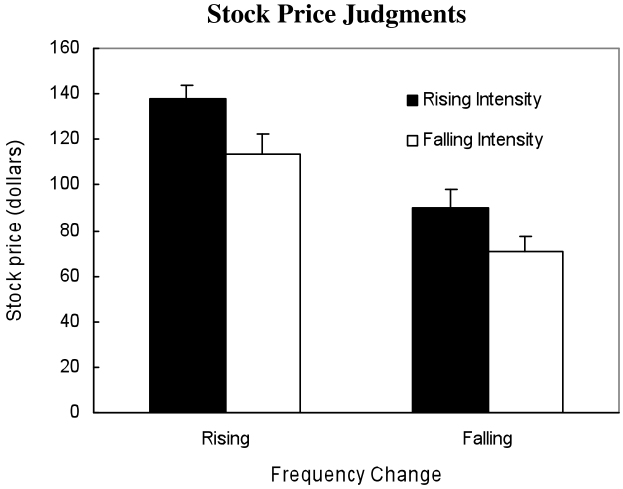

Figure 4.2: Perceptual interaction in auditory display. In a stock market sonification where terminal stock price was mapped to frequency change and the number of shares traded was mapped to intensity change, listeners gave different terminal price estimates for the same amount of frequency change depending on whether the concurrent intensity rose or fell. When frequency and intensity both rose, prices were judged to be higher than when frequency rose the same amount, but intensity fell. *(Adapted from Neuhoff, Kramer & Wayand, 2002).*

It is not at all uncommon in the context of sonification and auditory display for display designers to use various auditory dimensions to represent distinct variables in a data set. For example, when sonifying historical weather patterns, a display designer might use pitch change to represent the change in temperature, loudness change to represent the changes in the amount of precipitation, and timbre change to represent the relative change in humidity. Changes in the three weather variables can be easily represented by changes in three separate physical characteristics of the acoustic signal (frequency, amplitude, and spectrum). However, the perceptual interaction that occurs can be problematic. Although loudness is *principally* determined by the amplitude of a sound, there is good evidence that loudness is also more subtly influenced by the frequency (or pitch) of a sound. Fletcher & Munson (1933) were the first to show that the loudness of pure tones of equal intensity varied as a function of frequency. Their “equal-loudness contours” showed that listeners are most sensitive to sounds between 2 k kHz and 5 kHz. Similarly, Stevens (1935) showed that intensity can influence pitch. His “equal-pitch contours” showed that tones that differ in intensity can differ in frequency by up to 3% and still be perceived as equal in pitch. Timbre can interact with both pitch and loudness in similar ways.

Neuhoff, Kramer, & Wayand (2002) showed that pitch and loudness interact in auditory displays (see Figure 4.2). In a sonification of fictional stock market data, changes in stock price were mapped to changes in pitch, and changes in the number of shares traded were mapped to loudness. Rising pitch represented an increase in the price of a stock, and rising loudness represented an increase in the number of shares traded. In two contrasting conditions, listeners judged the terminal price of a stock. In one condition, the stock price rose while the number of shares also rose. This was represented with a sound that increased in both pitch and loudness. In the other condition, the stock price also rose (by the same amount as in the first condition). However, as the pitch increased to represent the rising price of the stock, the number of shares traded fell, thus loudness *decreased*. Despite the fact that the terminal pitch in each condition was the same and the stock price should be perceived as the same in each condition, listeners judged the price to be higher when both pitch and loudness rose than when pitch rose and loudness fell. Similarly listeners rated the price as *lower* when pitch and loudness both fell than when pitch fell and loudness rose. In other words, when the two dimensions changed in the same direction, the amount of change in one dimension was perceived as greater than when they changed in opposite directions.

4.2.6 Advantageous Effects of Dimensional Interaction

Auditory dimensions can be detrimental when separate variables are mapped to different auditory dimensions. However, there are cases when the interaction of auditory dimensions can be advantageously used in an auditory display. Mapping a single variable to multiple auditory dimensions has been shown to make the changes in that variable more salient than mapping it to single dimensions alone. For example, in sonifying changes in the volume of internet traffic on a particular site, one might use changes in loudness to denote changes in the amount of traffic, with higher loudness representing a higher volume of traffic. However, the change in traffic would be more perceptually salient if it were redundantly mapped to more than one dimension. Hansen and Ruben (2001) represented an increase in traffic by mapping it loudness, timbre, and repetition rate of a tone. So, an increase in traffic would yield a tone that gets brighter in timber, repeated faster, and also gets louder. This kind of “redundancy mapping” is effective in situations where absolute values in data are of secondary importance to changes and trends.

Redundancy mapping is also useful in auditory process monitoring tasks, particularly during “eyes busy” situations. Peres and Lane (2005) showed that redundant pitch and loudness mapping improved performance in a situation where listeners had to monitor auditory box plots while simultaneously performing a visual task. Importantly, the gains in performance due to redundancy mapping only occurred for auditory dimensions that have been shown to interact or are considered “integral” (such as pitch and loudness). When “separable” auditory dimensions (e.g., pitch and tempo) were mapped redundantly performance was not improved over the case in which only a single auditory dimension was used.

4.3 Auditory-Visual Interaction

There is a history in perceptual research of greater research efforts toward vision than audition, and a concentration on a single modality rather than on how vision and audition interact. We have relatively detailed accounts of the function of structures in the visual pathways when compared with those in audition. We know even less about the physiological interaction of the two systems. However, there are some clear examples of auditory and visual interaction at both the neurological and behavioral levels that have important implications for auditory display.

Perhaps the most famous example of auditory-visual interaction comes in the area of speech perception. The "McGurk Effect" occurs when visual and auditory speech tokens are mismatched, but presented simultaneously. For example, subjects may be presented with a video of a talker saying the syllable /ba/ with an accompanying audio track that says /ga/. In this case, listeners overwhelmingly report hearing the syllable /*da*/ (see video example **S4.1**). The work provides strong evidence for multimodal speech perception as does work showing that speech intelligibility increases when subjects can both hear and see the talker (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Sumby & Pollack, 1954) Thus, although the nature of auditory displays are such that they are most useful in "eyes busy" or low vision conditions, auditory displays that incorporate speech might benefit from the use of video to increase to reliability of the display if the conditions warrant.

Localization is another area in which strong auditory-visual interaction has been found. Visual performance in localization tasks is generally better than auditory performance. However, when subjects can use both their eyes and ears to localize an object, performance outpaces that which occurs in the visual only condition (Spence, 2007). The interdependence of vision and audition are particularly important in displays that require spatial estimates of a target that is both auditory and visual. In some contexts if the auditory and visual signals emanate from different locations, "visual capture" (or the "ventriloquist effect") will occur and users can perceive the audio signal as emanating from the location of the visual signal. However, in other contexts, the target can be perceived as somewhere in between the two signals (Alais, & Burr, 2004; Pick, Warren & Hay, 1969). This suggests a cognitive representation of external space that may be invariant across perceptual modalities. This arrangement allows, for example, that an auditory object that is heard but not seen can be spatially referenced with an object that is seen but not heard.

4.4 Auditory Space and Virtual Environments

The details of "how" we are able to perceive auditory space and motion are covered in the previous chapter. This section examines how this ability can be leveraged for use in auditory displays in both real and virtual environments and how the perception of auditory space interacts with vision.

Despite remarkable human ability to localize sound sources, the spatial resolution of the auditory system pales in comparison to what we can resolve visually. This, along with other visual advantages may contribute to the notion that humans are primarily "visual" beings. However, our strong reliance on vision may actually overshadow the degree to which we do rely on our ears for spatial localization and navigation in the real world. For example, there are some particular advantages that are obtained when localizing objects with our ears that cannot be obtained when localizing objects with our eyes. Obviously, the perception of auditory space does not require light. So, darkness and other poor viewing conditions do not present any great difficulty for auditory localization tasks. We can also hear objects that are hidden or occluded by other objects. Finally, while the field of vision is limited to approximately 120 degrees in front of the viewer, listeners can detect sounding objects 360 degrees around the head. Chronicling the advantages and disadvantages of the two systems might lead one to think that they are somehow in competition. However, the two systems work seamlessly together, each with strengths compensating for deficits in the other's repertoire of localization abilities. The result is an integrated multi-modal localization system that has evolved to help us localize objects and navigate a complex environment. As virtual environments become more common, our knowledge of both auditory and visual spatial perception will be crucial in creating environments that maintain a sense of presence.

In some environments where spatial auditory display is employed, the interaction of vision and audition can be of critical concern to display designers. For example, the spatial coincidence of auditory and visual representations of an object in a display will increase the sense of presence as well as the overall localization accuracy of the user. In other cases (e.g., auditory displays for the visually impaired) the focus on auditory-visual spatial interaction is decidedly less important. Nonetheless, the use of spatial information in auditory display is increasing. Advances in technology and our knowledge of how the auditory system processes spatial information has led to the emergence of virtual environments that realistically recreate 3-dimensional spatial auditory perception.

Many of these virtual auditory displays use Head Related Transfer Functions (HRTFs) to present binaural acoustic signals over headphones that mimic how the sounds would be received in a natural environment. The acoustic cues to spatial location (see chapter 3, this volume) can be manipulated as the user moves through a virtual environment. For example, a sound presented to the right of a listener will be perceived as louder in the right ear than in the left. However, when the listener's head turns to face the sound, a head-tracking device detects the movement of the head. The system detects the change in head position and the rendering system adjusts the level of the sound to be equal in the two ears, now equidistant from the source. All of the other cues to localization (e.g., interaural time differences and pinnae cues) are adjusted in a similar manner. Other systems use an array of loudspeakers that surround the listener. There are advantages and disadvantages to both approaches. However, the goal of both types of systems is to render the acoustic properties of the sound source and the environment such that the listener experiences the sound as though they were listening in the environment in which the sound would normally occur (Lokki, Savioja, Vaanaanaen, Huopaniemi, & Takala, 2002). When done well, the result is a spatial auditory display that in almost every way is more realistic and has better resolution than current visual virtual environments. Virtual auditory environments have applications in many domains. In addition to providing highly controlled environments in which researchers can study the psychology and physiology of auditory spatial perception (e.g., Nager, Dethlefsen, Münte, 2008), virtual auditory displays are used in psychiatry, aviation, entertainment, the military, as aids for the visually impaired, and in many other areas. The addition of spatialized sound in virtual environments does not only add to the auditory experience. It also increases the overall sense of presence and immersion in the environment (Hendrix & Barfield, 1996; Viaud-Delmon, Warusfel, Seguelas, Rio, & Jouvent, 2006). Spatial coherence between visual and auditory objects in such environments are crucial to maintaining this sense of presence for the user.

Navigation performance in virtual environments is significantly better when spatial auditory information is present than when it is not (Grohn, Lokki, & Takala, 2003), and researchers have also shown that auditory localization performance for some listeners is comparable in real and virtual auditory environments (Loomis, Hebert, & Cicinelli, 1990). These latter findings are particularly encouraging to those involved in developing displays that are used for navigation (e.g., Seki & Sato, 2011).

4.5 Space as a Dimension for Data Representation

Virtual auditory environments provide a host of new possibilities for auditory display. The ability to use space as another dimension creates interesting possibilities for sonification and auditory display designers. Recent advances in technology have caused a growth in the use of spatialised sound in the areas of sonification and auditory display.

In one interesting example, Brungart and colleagues (2008) designed an auditory display for pilots that represented the attitude of the aircraft relative to the horizon (i.e. the plane's pitch and roll). Changes in roll were represented spatially by moving an audio signal back and forth as necessary between the left and right headphones. When the plane banked to the left, the signal moved to the right ear and vice versa. Additionally, changes in the plane's pitch (relative "nose-up" or "nose-down" position) were represented by a spectral filtering process. When the plane was nose-up, a spatially diffuse and low pitched characteristic was present in the stimulus, indicating that the nose of the aircraft should brought down to a more level flight position. When the aircraft was "nose-down" the signal was changed to a high pitched characteristic indicating that the nose of the plane should be pulled up. Straight and level flight was indicated by a spectrally unchanged signal that was equally centered between the right and left headphones. Importantly, the audio signal that was fed into the system could be anything, including music selected by the pilots. This technique has the advantage of greatly reduced annoyance and listener fatigue as well as higher compliance (i.e. the willingness of pilots to use the system).

Auditory spatial cueing has also been shown to be effective in automotive applications. Ho and Spence (2005) designed a display in which spatial auditory warnings facilitated visual attention in the direction of the auditory warning. Moreover, performance of emergency driving maneuvers such as braking or acceleration was improved by the use of spatial auditory displays.

4.6 Rhythm and Time as Dimensions for Auditory Display

One indication of the dominance of vision over other senses in humans is the tremendous disparity in the amount of cortex devoted to visual processing when compared to the other sensory modalities. Thus, as one might expect, the visual system tends to show better performance on many types of perceptual tasks when compared to audition (e.g., spatial localization). However, when it comes to rhythmic perception and temporal resolution, the auditory system tends to perform significantly better than the visual system. Thus, auditory display is particularly well suited for domains in which rhythmic perception and temporal discrimination are critical and domains in which the underlying data lend themselves to rhythmic and temporal variation, particularly when the rate of presentation is within the optimal sensitivity range of tempi for the user (Jones, 1976; Jones & Boltz, 1989).

Sound is inherently temporal, and differences in the timing and tempo of acoustic information has been studied extensively. Differences in tempo between displays and changes in tempo within a single display can be used effectively to convey relevant information. For example, urgency is generally perceived as higher when acoustic stimuli are presented at faster rates (Edworthy, Loxley, & Dennis, 1991; Langlois, Suied, Lageat & Charbonneau, 2008). Changes in tempo can also be used to indicate directional information (e.g., “up”, “down”) although tempo as an indicator of direction may be a relatively weak acoustic cue when compared to similar changes in pitch and loudness (Pirhonen & Palomäki, 2008). Semantic meaning of fast and slow rhythmic tempi has even been examined in the context of earcons (Palomäki, 2006).

Sensitivity to rhythm can also be exploited to indicate processes or data anomalies. For example, Baier, Hermann, and Stephani (2007; Baier & Herman, 2004) used variation in rhythm to indicate differences between epileptic and non-epileptic activity in human EEG data. Changes in rhythm and tempo have been also used in biofeedback systems designed for stroke rehabilitation (Wallis, et al. 2007).

Changes in rhythm and tempo can be used to indicate changes in the state or value of sonified data. However, simply speeding up the auditory display can also yield display benefits, particularly in displays that use speech. Listeners can retain a good deal of intelligibility even when speech is presented up to three times its normal rate (Janse, Nooteboom & Quené, 2003). The ability to perceive speech at faster rates than it is normally produced has been explored in a wide array of applications that range from screen readers for the visually impaired, to complex communication systems, to mobile phones. For example, many complex workstations feature simultaneous voice communication systems. Intelligibility in multiple talker systems generally decreases as the number of talkers goes up. Successful efforts to increase multi-talker intelligibility have generally focused on spatializing the talkers such that each voice emanates from a different position in space relative to the listeners (e.g., Brungart & Simpson, 2002). However, recent methods have also employed the dimension of time.

Brock, et al (2008) showed that in a four talker situation, speech that was artificially sped up by 75% and presented serially was understood significantly better than speech presented at normal speeds concurrently. Walker, Nance, and Lindsay (2006) showed that extremely fast speech can improve navigation through auditory menus in a cell phone application. As opposed to earcons or auditory iconsthe fast speech or “spearcons” yielded faster and more accurate user performance. Spearcons are produced by speeding up text-to-speech audio output until it is no longer perceived as speech. However, the spearcon still retains some similarity to the original speech signal from which it was derived. The time required to learn an auditory menu also appears to be reduced when the menu is presented with spearcons rather than earcons (Palladino & Walker, 2007). The majority of the research on spearcons has been conducted from the perspective of improving human performance in auditory display settings. Thus, little is known about the underlying cognitive mechanisms that afford the enhanced performance.

4.7 Auditory Scene Analysis

Sounds generally do not occur in isolation. At any given instant numerous sound sources create separate acoustic waves that can reach the ear simultaneously. In fact, in most environments it is unusual to hear a single sound in isolation. The hum of a computer fan is heard simultaneously with the ticking of a clock, a conversation in the next room, and the muffled sound of passing cars on the roadway outside. When the acoustic waves from these various sound sources reach the tympanic membrane (eardrum) they result in a single highly complex mechanical signal that, were it examined visually via sonograph, might appear to be almost random. Yet, perceptually listeners have little difficulty extracting the source information from this complex signal and can easily hear distinct sound sources. In other words, it is rarely difficult to tell where one sound stops and another begins. The process of segregating these auditory sources is referred to as *auditory scene analysis*, and the ease with which we accomplish the task belies its tremendous complexity.

Consider for example, the many technologies that now respond to voice commands. Cell phones, computers, automobiles, and many other devices can decipher simple voice commands and produce a requested action, provided there are no significant competing background sounds. Have just two or three people speak to a voice activated device simultaneously and the device fails to detect where one voice ends and another begins, a task that human listeners can do with relative ease. Thus, despite our tremendous technological advances, we have yet to develop voice-activated technology that might work well, for example, at a cocktail party (Cherry, 1953).

Al Bregman pioneered the study of auditory scene analysis by asking questions about audition that at the time were considered “non-traditional”. His Ph.D. in cognitive psychology was completed at Yale in a laboratory that primarily studied vision. He subsequently pursued his interest in auditory perception undertaking research that was a strong departure from the traditional work being performed in psychoacoustics at the time. He applied many of the same techniques and questions that were being asked about visual phenomena to auditory phenomena. Because of his work, people now speak regularly of “auditory objects” and “auditory streams”.

The perceptual work on auditory scene analysis has important implications for auditory display. The fact that listeners can attend to separate sources or auditory streams allows sonification designers to exploit our auditory perceptual organization abilities and simultaneously present distinct aspects of a multidimensional data set in distinct auditory streams.

Acoustic characteristics and attentional factors can both influence how the auditory system perceptually organizes the auditory scene. At the acoustic level, individual sound sources have certain acoustic regularities in the sounds that they produce that can be used by the auditory system to parse an auditory stream. For example, sounds that are similar in frequency are more likely to be allocated to the same source. Thus, in sonifying a multidimensional data set it is common to separate different variables by differences in frequency range. One variable might be represented by low frequency sounds and another by high frequency sounds. The separation in frequency of the two “streams” makes it more likely that the two variables being sonified will also remain independent. Similarly, sounds that are similar in timbre are more likely to be perceptually grouped together. Thus, a common technique is to use different musical instruments to represent separate aspects of the underlying data.

Pitch and timbre differences can be manipulated independently to affect how auditory grouping occurs. For example, the grouping effect that occurs by making sounds similar in pitch can be counteracted by making them dissimilar in timbre and vice versa (Singh, 1987). Conversely, the grouping effect can be made stronger by using redundant segregation cues in each stream. For example, the differences in both pitch and timbre between a bass guitar and a piccolo would provide better stream segregation than only the timbre differences of say, a saxophone and trumpet played in the same frequency range.

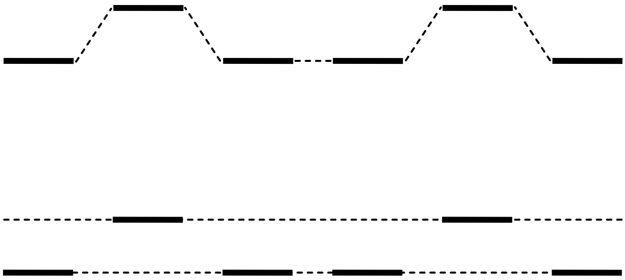

Figure 4.3: Alternating high and low pitched tones can either be perceived as one or two auditory streams depending on presentation rate and the distance in pitch between the tones.

Differences in other acoustic characteristics such as loudness and spatial location can also be used to parse sound sources. Although loudness level may not be as strong a cue to grouping as other acoustic characteristics, sounds presented at similar levels nonetheless tend to group together (Hartmann & Johnson, 1991; Van Noorden, 1975). Spatial location is a strong cue to auditory stream segregation. Sounds that come from the same location tend to be grouped together. A lack of spatial coherence often prevents sounds from being perceptually grouped together. For example, a sequence of tones presented to alternating ears tends not to form a single auditory stream (Van Noorden, 1975). Spatial separation of sources using binaural cues to localization is a particularly effective means for segregating real world sources such as multiple talkers (Hawley, Litovsky, & Culling, 2004).

Tempo and rhythm also interact with auditory stream segregation. However, rather than parsing simultaneous sounds into distinct auditory objects, tempo and rhythm effects are more likely to occur with sequentially presented stimuli. Van Noorden (1975) presented listeners with alternating high and low pitched notes that could either be perceived as one or two separate streams (see Figure 4.3). When perceived as a single stream the notes are heard as a galloping rhythm that goes up and down in pitch. When perceived as two streams the notes are heard as two repeating patterns each with a regular isochronous rhythm. The tempo at which the stimuli are presented can influence whether the notes are perceived as one stream or two with faster tempi being more likely to induce the perception of two streams. Moreover, cues which aid in simultaneous stream segregation can also influence sequential segregation (Micheyl, Hunter & Oxenham, 2010). For example, the amount of separation in frequency between the notes can influence how the streams are perceived. Greater frequency separation makes it more likely that the notes will be perceived as two streams (Bregman, 1990).

In addition to the lower level acoustic characteristics of sound sources, attention and higher order cognitive processes can affect how the auditory scene is parsed (Bregman, 1990; Carlyon, Cusack, Foxton, & Robertson, 2001; Snyder & Alain, 2007; Sussman & Steinschneider, 2009). Prior knowledge, expectations, selective attention, and expertise can all influence the landscape of the auditory scene. These cognitive processes work in concert with the acoustic characteristics when listeners parse auditory objects (Alain, Arnott, & Picton, 2001).

4.8 Auditory Cognition

There is a rich history of psychoacoustic research on the "sensory" aspects of audition. Conversely, "auditory cognition" has received comparatively little attention. Incoming acoustic information is transformed into a neural signal at the level of specialized cells in the inner ear. With the exception of speech and music, this is where the study of audition often stopped. However, in addition to the incoming acoustic signal that arrives at the eardrum, the listener's prior knowledge, experience, expertise, and expectations can all influence how acoustic information is perceived. Cognitive psychologists have come to call these kinds of effects "top-down" processing to distinguish them from the "bottom-up" processing that occurs when acoustic information is received, transformed into a sensory signal, and passed "up" to higher cortical areas. The effects of top-down processing are widespread (though perhaps not well known) in auditory display environments. Any type of effect in user performance due to the expertise of the user, training, or the expectations of the user comes under the umbrella of top-down effects (Strait, Kraus, Parbery-Clark, & Ashley, 2010; Sussman, Winkler, & Schröger, 2003).

An example of top-down cognitive processing occurs in a phenomenon called the "phonemic restoration effect". In natural listening environments speech sounds are often briefly interrupted or masked by other environmental sounds. Yet, this rarely interferes with the listener's comprehension of the message. Warren (1970) showed that if a phoneme (i.e, the smallest segment of a word that still imparts meaning) is removed from a word and replaced with noise or a cough, listeners still hear the missing phoneme. Moreover, they have great difficulty even indicating where the cough or noise occurred in the utterance. The effect has been rigorously researched and is the result of top-down perceptual processing (Samuel, 2001).

The simple act of recognizing a friend's familiar voice also requires top-down processing. Subsequent to the transformation of the acoustic signal into a neural impulse, the stimulus must be identified as a voice, likely engaging many of the mechanisms that process the various aspects speech, including syntax, semantics, and even emotion. Memory must be activated, and the incoming signal matched to a cognitive representation of your friend's voice. All of this occurs in an instant, and you can then recognize that your friend is talking to you, he wants to get something to eat, and he sounds a little sad. The prior experience, memory, and expectations of the listener can shape the perception of sound. Similar processes must occur for non-speech sounds. Recognizing and responding appropriately to the sound of a car horn, a baby's cry, or gunfire can have life or death implications.

Although researchers are beginning to make progress in understanding some of the complex processes that occur in "auditory meaning making" for speech, they are not yet completely understood. When it comes to understanding the cognitive processes of the non-speech sounds typically used in auditory display, we know even less. Thus, in order to understand sound and derive real world meaning from these neural signals, a more thorough investigation is required. Cognition and action in response to auditory stimuli are crucial not only in auditory display environments, but in almost all real world situations.

4.8.1 Cognitive Auditory Representations

Cognitive or "mental" representations of stimuli have a rich history in cognitive psychology. They are also a potentially fruitful area for designers of auditory displays. The idea that a cognitive representation of an external stimulus could even exist was at one time quite controversial, and the specifics of such representations are still debated among psychologists and cognitive scientists. There is clearly subjective or anecdotal evidence of cognitive representations. When asked, for example, to imagine their kitchen, most people can bring a visual image of their kitchen to mind and describe it in some detail. From an experimental perspective, behavioral and neuroimaging studies have provided rather convincing evidence that the brain does store some kind of representation of stimuli from the external world.

In the auditory domain, there is also evidence for cognitive representations of acoustic stimuli. As in the visual domain, there is abundant subjective and anecdotal evidence. Almost anyone will admit to being able to imagine the sound of a car horn, a bird chirping, or of eggs frying in a pan. There is also abundant experimental evidence for "auditory imagery". In one ingenious study by Halpern and Zatorre (1999), subjects listened to simple melodies while connected to a Positron Emission Tomography (PET) scanner. The PET scanner allows researchers to identify areas of brain activation during various activities or when various stimuli are presented. In one condition the subjects were simply asked to listen to the song. In another condition subjects were played only the first half of the song and asked imagine the rest by "singing it in their head". The surprising finding was that the same areas of the brain were active during the silent "imagined" portion of the song as were active when the song was actually heard. This work suggests that auditory "cognitive representations" may in fact simply be the occurrence of a pattern of neural firing in the absence of a stimulus that would occur if the stimuli were actually present.

Surprisingly, cognitive representations of real world sounds have not been widely used by sonification designers as a means of representing variable data sets. The majority use simple changes in pitch, loudness or timbre to represent changes in the variables of interest. The result is often a changing auditory signal that has no direct cognitive representation of the underlying data for the listener. This is certainly not to say that associations between the changing acoustic characteristics and the data set cannot be learned; only that it is a secondary process to understand, for example, that a change in timbre represents a change in temperature. Moreover, when multivariate datasets are sonified, simultaneous changes in pitch, loudness, and timbre are commonly used in a single signal to represent various changes in data. However, the underlying data in this example are subject to distortions from the perceptual interaction effects outlined above.

An alternative to this sonification technique has been proposed that involves mapping changes in real world auditory events to changes in the underlying data set. Gaver (1993) suggested that listeners attend to "auditory events" in a way that makes the physical characteristics of the sound source an important factor in auditory perception of non-speech sounds. So, rather than hearing "... *a quasi-harmonic tone lasting approximately three seconds with smooth variations in the fundamental frequency and the overall amplitude...*", listeners will report instead that they heard "*A single-engine propeller plane flying past*", (Gaver, 1993, p. 285–286). The upshot is that listeners consciously process events, not acoustics.

Neuhoff and Heller (2005) suggested that this "event based" representation might be effectively used in sonification. For example, rather than mapping increasing pitch to an increase in the data, a designer might instead map changes in the data to the pace of a real world auditory event that listeners are highly skilled at perceiving, such as footsteps (Li, Logan, & Pastore, 1991; Visell, et al., 2009). The advantage to this approach is twofold. First, the changes in these complex stimulus dimensions tend to be more familiar and easier to identify than changes in simple acoustic dimensions. Music novices, for example, often have difficulty describing pitch change as going "up" or "down" because they have not been had the necessary exposure to know that increases in frequency are related to "higher" pitch (Neuhoff, Knight & Wayand, 2002). However, most listeners can easily distinguish between fast and slow footsteps. Second, the problem of unwanted interacting perceptual dimensions can be avoided by using real world auditory events to represent changes in data. For example, if walking speed were used to represent one variable in a multivariate data set, the hardness of the surface might be used to represent another variable. Most listeners can identify specific properties of walking surfaces in addition to characteristics of the walker such as gender and height (Visell, Fontana, Giordano, Nordahl, Serafin& Bresin, 2009). The complexity of such an acoustic representation would yield large benefits in the simplicity of the perceptual interpretation of the data (Neuhoff & Heller, 2005).

4.8.2 Music and Data Representation

Perhaps some of the most structured auditory cognitive representations that exist are musical systems. Musical scales provide a formal structure or framework that can be leveraged in the design of effective auditory displays (Krumhansl, 1982; Jordan & Shepard, 1987; Shepard, 1982). Thus, given that one of the main goals of auditory display is to communicate information, auditory display can be informed by music theory. Rather than mapping data to arbitrary changes in frequency, many auditory displays map changes in data to changes in pitch that are constrained to standard culturally specific musical scales. For example, Vickers and Alty (1997; 2002; 2003) have employed melodic motifs to aid computer programmers in debugging code and to provide other programming feedback. Valenzuela (1998) used melodic information to provide users with integrity evaluation information about concrete and masonry structures. Melodic information in auditory display has even been used as a tool for mathematics instruction with middle school and high school students (Upson, 2002).

An advantage of using musical scales in sonification is that they may be perceived as more pleasant and less annoying than frequency change that is not constrained to musical scales. Although there has been ample work to show that differing levels of musical expertise can influence perceptual performance in a musical setting (e.g., Bailes. 2010), these differences can be minimized when the stimuli are interpreted in units that reflect the underlying data dimensions (Neuhoff, Knight, & Wayand, 2002). The effects of musical expertise on the perception of auditory displays have not been thoroughly investigated. Part of the difficulty in this area has been the lack of a well designed system for measuring musical expertise (Edwards, Challis, Hankinson & Pirie, 2000). Although there are tests of musical ability among musicians, there are few validated ways of examining musical ability among those who have no formal training in music (however, for one promising method see Ollen, 2006).

4.8.3 Perception and Action

The idea that our actions and the motor system are involved in perceiving the external world dates back to at least the late 1960s. Liberman and colleagues (1967) proposed that the speech signal is decoded in part by referring incoming speech sounds to the neuro-muscular processes that are used to produce them. In essence, we understand speech through the motor commands that are employed when we ourselves speak. The details of the "Motor Theory" of speech perception have been sharply debated over the years, but there are few who would doubt that perception and action are closely linked in many domains.

Advances in neuroimaging have yielded numerous investigations which show that regions of the brain that are responsible for motor activity are recruited to process incoming auditory stimuli, even when those stimuli are non-speech sounds. For example, Chen and colleagues (2008) showed that motor areas were active when subjects listened to a rhythmic pattern in anticipation of tapping along with the rhythm later. Even when subjects were simply asked to listen to the rhythms with no knowledge that they would be asked to tap along later, the same motor regions were active. Similarly, pianists show activation in motor areas when simply listening to a piano performance (Haueisen, Knösche, 2001; Bangert, et al, 2006). The perception-action link is further evidenced by the finding that non-pianists (who presumably would not have the motor plans for a piano performance) do not show activation in motor areas when presented with the same music.

In another study, subjects were presented with "action sounds" that were consistent with human motor behavior (e.g., crunching, opening a zipper, crushing an aluminum can) and "non-action" sounds that did not require any motor behavior (e.g., waves on a beach, a passing train, or wind). The motor areas of the brain activated when the "action sounds" were presented were the same ones activated when the subjects actually performed the actions depicted in the sounds. However, motor areas were not recruited when listeners were presented with the non-action sounds. In addition to processing incoming stimuli, these so called auditory "mirror neurons" may be involved in facilitating communication and simulation of action (Kohler, Simpson l., 2002).

An important point taken from these studies is that the articulatory gestures that are used to produce "action sounds" may be as important as the acoustic structure of the sounds themselves. In other words, the link between the auditory and motor system appears to capitalize on the knowledge of the actions used to produce the sounds as much as the specific acoustic attributes per se. Thus, the use of real world sounds in auditory display discussed previously may tap into perceptual and "meaning making" processes that cannot be accessed with sounds that are more artificial. An additional distinction among real world sounds has been made by Giordano, McDonnel, and McAdams (2010). They used a sound sorting task with "living sounds" and "non-living sounds" and found that listeners differentiate non-living action and non-action sounds with an iconic strategy that does indeed focus on acoustic characteristics of the sound. The evaluation of living sounds, on the other hand, relied much more on a symbolic cognitive representation of the sound referent.

From the perspective of designing auditory displays, these findings suggest that the judicial use of environmental sounds rather than simpler artificial sounds might provide a better means of communicating the information to be displayed. Millions of years of evolution have produced neural and cognitive architecture that is highly sensitive to meaningful real world environmental sounds. Perceptual processing of these sounds appears to happen in a way that is fundamentally different from that which occurs with simple arbitrary beeps and buzzes. We know that simply mapping a sound that has a clear environmental referent (i.e. auditory icons see chapter 13) to a particular display dimension increases user response time and accuracy in the display over more arbitrary mappings (McKeown & Isherwood, 2007). Future research may demonstrate even greater gains with environmental sounds have a clear *behavioral* referent which maps to a specific motor action.

4.9 Summary

The ease with which we perceive the auditory world masks the complexity of the process of transforming acoustic waves into meaning and responsive behavior. Basic acoustic dimensions such as pitch, loudness and timbre can be used to represent various aspects of multidimensional data. However, extreme care and an intentional approach should be taken in understanding the perceptual interactions that occur with these kinds of dimensions. Auditory perception acts in concert with other sensory modalities, and cross modal influences with vision and other senses can influence perception and performance in an auditory display. Higher order acoustic characteristics, including time, and space, are also common vehicles through which acoustic information is used to represent data. These factors interact with the cognitive processes involved in auditory scene analysis, music and speech, and perception-action relationships to form a complex foundation upon which effective auditory displays can be designed.

Bibliography

[1] Alain C., Arnott S. R., Picton T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. J Exp Psychol Hum Percept Perform. 27(5):1072–89.

[2] Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Current Biology, 14, 257–262.

[3] Anderson, J., & Sanderson, P. (2009). Sonification design for complex work domains: Dimensions and distractors. Journal of Experimental Psychology: Applied. 15(3), 183–198.

[4] Baier, G. & Hermann, T. (2004). The sonification of rhythms in human electroencephalogram. Proceedings of the International Conference on Auditory Display.

[5] Baier, G., Hermann T., & Stephani U. (2007). Event-based sonification of EEG rhythms in real time. Clinical Neurophysiology. 1377–1386.

[6] Bailes, F. (2010). Dynamic melody recognition: Distinctiveness and the role of musical expertise. Memory & Cognition 38(5), 641–650.

[7] Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H, Heinze, H.J., & Altenmüllera, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. NeuroImage, 30, 917–926.

[8] Bregman, A. S. 1990. Auditory Scene Analysis. MIT Press, Cambridge, MA.

[9] Brock, D., McClimens, B., Wasylyshyn, C., Trafton, J. G., & McCurry, M. (2008). Evaluating listeners'

attention to and comprehension of spatialized concurrent and serial talkers at normal and a synthetically faster rate of speech. Proceedings of the International Conference on Auditory Display.

[10] Brungart D. S. & Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. J. Acoust. Soc. 112, 664–676.

[11] Brungart D. S., & Simpson, B. D., (2008). Design, validation, and in-flight evaluation of an auditory attitude ondicator based on pilot-selected music. Proceedings of the International Conference on Auditory Display.

[12] Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. Journal of Experimental Psychology: Human Perception and Performance, 27, 115–127.

[13] Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008) Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. Journal of Cognitive Neuroscience, 20(2):226–239.

[14] Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. Journal of Acoustic Society of America 25 (5): 975–979.

[15] Edwards, A. D. N., Challis, B. P., Hankinson, J. C. K., & Pirie, F. L. (2000). Development of a standard test of musical ability for participants in auditory interface testing. Proceedings of the International Conference on Auditory Display.

[16] Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design – relationship between warning sound parameters and perceived urgency. Human Factors, 33(2), 205–231.

[17] Fletcher, H. & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. Journal of the Acoustical Society of America, 5:82–108.

[18] Flowers, H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions, Proceedings of the International Conference on Auditory Display, 406–409.

[19] Flowers, J. H.,Whitwer, L. E., Grafel, D. C., & Kotan, C. A. (2001). Sonification of daily weather records: Issues of perception, attention and memory in design choices. Proceedings of the International Conference on Auditory Display.

[20] Garner, W. R. (1974). The Processing of Information and Structure. Potomac, MD: Erlbaum.

[21] Gaver, W. W. (1993). How do we hear in the world? Explorations of ecological acoustics. Ecological Psychology, 5(4): 285–313.

[22] Giordano, B. L., McDonnell, J., & McAdams, S. (2010). Hearing living symbols and nonliving icons: Category-specificities in the cognitive processing of environmental sounds. Brain & Cognition, 73, 7–19.

[23] Gröhn, M., Lokki, T., & Takala, T. (2003). Comparison of auditory, visual, and audio-visual navigation in a 3D space. Proceedings of the International Conference on Auditory Display, 200–203.

[24] Halpern, A. R. & R. J. Zatorre. 1999. When that tune runs through your head: an PET investigation of auditory imagery for familiar melodies. Cerebral Cortex 9: 697–704.

[25] Hansen, M. H. & Rubin, B (2001). Babble online: applying statistics and design to sonify the internet. Proceedings of the International Conference on Auditory Display.

[26] Hartmann, W. M., and Johnson, D. (1991). Stream segregation and peripheral channeling. Music Perception 9, 155–183.

[27] Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. Journal of Cognitive Neuroscience, 13(6), 786–79.

[28] Hawley, M. L., Litovsky, R. Y., & Culling, J. F. 2004. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. Journal of the Acoustical Society of America, 115, 833–843.

[29] Hendrix, C. & Barfield, W. (1996). The sense of presence within auditory virtual environments. Presence: Teleoperators and Virtual Environments. 5, (3), 290–301.

[30] Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of Experimental Psychology: Applied, 11, 157–174.

[31] Janata, P., & Childs, E. (2004). MarketBuzz: Sonification of real-time financial data. Proceedings of The 10th

Meeting of the International Conference on Auditory Display.

[32] Janse, E., Nooteboom, S. & Quené, H. (2003). Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. Speech Communication 41, 287–301.

[33] Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review 83: 323–355.

[34] Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. Psychological Review 96: 459–491.

[35] Jordan, D. S. & Shepard, R. N. (1987). Tonal schemas: Evidence obtained by probing distorted musical scales. Special Issue: The understanding of melody and rhythm, 41, 489–504.

[36] Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V. and Rizzolatti, G. (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science, 297: 846–848.

[37] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N.; Neuhoff, J., Bargar, R., Barrass, S., Berger, J., Evreinov, G., Fitch, W., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., Shinn-Cunningham, B., Simoni, M., Tipei, S. (1999). The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM: ICAD.

[38] Krumhansl, C. L. (1983). Perceptual structures for tonal music.,“ Music Perception, 1, 28–62.

[39] Langlois, S., Suied, C., Lageat, T. & Charbonneau, A. (2008). Cross cultural study of auditory warnings. Proceedings of the International Conference on Auditory Display.

[40] Li, X.-F., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. Journal of the Acoustical Society of America, 90, 3036–3049.

[41] Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review 74 (6): 431–461.

[42] Lokki, T., Savioja, L., Väänänen, R., Huopaniemi, J., Takala, T. (2002). Creating interactive virtual auditory environments. IEEE Computer Graphics and Applications, special issue “Virtual Worlds, Real Sounds”, 22(4), 49–57.

[43] Loomis, J. M., Hebert, C., & Cicinelli, J. G. (1990). Active localization of virtual sounds. Journal of Acoustical Society of America, 88, 1757–1764.

[44] McKeown, D., & Isherwood, S. (2007). Mapping candidate within-vehicle auditory displays to their referents. Human Factors, 49(3), 417–428.

[45] Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. Perception and Psychophysics, 48, 169–178.

[46] Micheyl, C., Hunter, C., Oxenham, A. J. (2010). Auditory stream segregation and the perception of across-frequency synchrony. Journal of Experimental Psychology: Human Perception and Performance. 36(4), 1029–1039.

[47] Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head mov

[48] Nager W., Dethlefsen C., Münte T. F. (2008). Attention to human speakers in a virtual auditory environment: brain potential evidence. Brain Research, 1220, 164–170.

[49] Neuhoff, J. G. & Heller, L. M. (2005). One small step: Sound sources and events as the basis for auditory graphs. Proceedings of the 11th International Conference on Auditory Display.

[50] Neuhoff, J. G., Knight, R. & Wayand, J. (2002). Pitch change, sonification, and musical expertise: Which way is up? Proceedings of the International Conference on Auditory Display.

[51] Neuhoff, J. G., Kramer, G., & Wayand, J. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? Journal of Experimental Psychology: Applied. 8 (1), 17–25.

[52] Ollen, J. E. (2006). A criterion-related validity test of selected indicators of musical sophistication using expert ratings. Dissertation, Ohio State University.

[53] Palladino, D., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus

earcons. Proceedings of the International Conference on Auditory Display, 274–279.

[54] Palomäki, H. (2006). Meanings conveyed by simple auditory rhythms. Proceedings of theInternational Conference on Auditory Display, 99-104.

[55] Peres, S. C., & Lane, D. M. (2005) Auditory Graphs: The effects of redundant dimensions and divided attention, Proceedings of the International Conference on Auditory Display, 169–174.

[56] Pick, H. L., Warren, D. H., , & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. Perception & Psychophysics, 6(4), 203–205.

[57] Pirhonen, A., Palomäki, H. (2008). Sonification of Directional and Emotional Content: Description of Design Challenges. Proceedings of the International Conference on Auditory Display.

[58] Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. Psychological Science, 12, 348–351.

[59] Seki Y., Sato T. A. (2011) Training system of orientation and mobility for blind people using acoustic virtual reality. IEEE Trans Neural Syst Rehabil Eng. 1, 95–104.

[60] Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. Psychological Review, 89, 305–333.

[61] Singh, P. G. (1987). Perceptual organization of complex-tone sequences: a tradeoff between pitch and timbre? J Acoust Soc Am 82: 886–899.

[62] Smith, D. R. & Walker, B. N. (2002). Tick-marks, axes, and labels: The effects of adding context to auditory graphs. Proceedings of the International Conference on Auditory Display, 362–36

[63] Snyder, J. S., & Alain, C. (2007). Sequential auditory scene analysis is preserved in normal aging adults. Cerebral Cortex, 17, 501–5.

[64] Spence, C. (2007). Audiovisual multisensory integration. Acoustical Science and Technology, 28, 61–70.

[65] Stevens, S. S. (1935). The relation of pitch to intensity. Journal of the Acoustical Society of America, 6, 150–154.

[66] Stevens, S. S. & Davis, H. (1938). Hearing: Its psychology and physiology. Oxford, England: Wiley.

[67] Strait D. L., Kraus, N. Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. Hearing Research 261, 22–29.

[68] Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. Journal of the Acoustical Society of America, 26 (2) 212–215.

[69] Sussman E., Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. Neuropsychologia 47(3):771–785.

[70] Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. Psychonomic Bulletin & Review, 10(3), 630–637.

[71] Upson, R. (2002). Educational sonification exercises: Pathways for mathematics and musical achievement. International Conference on Auditory Display, Kyoto, Japan.

[72] Valenzuela, M. L. (1998). Use of synthesized sound for the auditory display of impact-echo signals: Design issues and psychological investigations. Dissertation Abstracts International: Section B: The Sciences and Engineering. Vol 58(12-B).

[73] van Noorden, (1975) Temporal Coherence in the Perception of Tone Sequences. Dissertation, Technical University Eindhoven.

[74] Viaud-Delmon I, Warusfel O, Seguelas A, Rio E, Jouvent R (2006). High sensitivity to multisensory conflicts in agoraphobia exhibited by virtual reality, European Psychiatry, 21(7), 501–508.

[75] Vickers, P. and Alty, J. L. (1996) CAITLIN: A Musical Program Auralisation Tool to Assist Novice Programmers with Debugging. Proceedings of the International Conference on Auditory Display.

[76] Vickers, P. and Alty, J. L. (2002) Using Music to Communicate Computing Information. Interacting with Computers, 14 (5). 435–456.

[77] Vickers, P. and Alty, J. L. (2003) Siren Songs and Swan Songs: Debugging with Music. Communications of

the ACM, 46 (7). 86–92.

[78] Visell, Y., Fontana, F., Giordano, B.L., Nordahl, R., Serafin, S., & Bresin, R. (2009). Sound design and perception in walking interactions. International Journal of Human-Computer Studies, 67 (11), 947–959.

[79] Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. Journal of Experimental Psychology: Applied, 8, 4, 211–221.

[80] Walker, B. N. (2007). Consistency of magnitude estimations with conceptual data dimensions used for sonification. Applied Cognitive Psychology, 21(5), 579–599.

[81] Walker, B. N., & Ehrenstein, A. (2000). Pitch and pitch change interact in auditory displays. Journal of Experimental Psychology: Applied, 6, 15–30.

[82] Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. Neuhoff (Ed.), Ecological Psychoacoustics (pp.150–175). New York: Academic Press.

[83] Walker, B. N., & Lane, D. M. (2001). Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. Proceedings of the International Conference on Auditory Display, 90–94.

[84] Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus. Proceedings of the International Conference on Auditory Display, 63–68.

[85] Wallis, I., Ingalls, T., Rikakis, T., Olsen, L., Chen, Y., Xu, W., & Sundaram, H. (2007). Real-time sonification of movement for an immersive stroke rehabilitation environment. Proceedings of the International Conference on Auditory Display, 497–503.

[86] Warren, R. M. (1970). Perceptual restoration of missing speech sounds. 167, 392–393.

[87] Wegner, K. (1998). Surgical navigation system and method using audio feedback. Proceedings of the International Conference on Auditory Display.

[88] Won, S. Y. (2005). Auditory display of genome data: Human chromosome 21. Proceedings of the International Conference on Auditory Display, 280–282.