Ecological Considerations for Cochlear Implant Research

Abstract

In traditional psychoacoustics, stimuli typically consist of short sequences of pure or complex tones, noise bursts, or simple combinations of these sounds. In contrast to typical laboratory stimuli, *everyday sounds* are inharmonic, noisy, non-stationary, and have irregular temporal patterns. Ecological Psychology maintains that the environment and an observer are inseparable, and that perception must be studied in real-life conditions. This paper presents ideas for the development of "new" stimuli that could prove invaluable for evaluating cochlear implant efficacy in a life-like but controlled laboratory environment. The motivation and development of these proposed stimuli incorporates ideas founded in Ecological Psychology.

Background: Ecological Psychology

Ecological psychology, founded by J. J. Gibson, is based on the theory that perception is *direct*; i.e., unmediated by inference or memory, and that animals and their environments are "inseparable pairs" (Gibson, 1966, 1979). Rather than internal representations of the world, animals detect and use information about the world as required. The animal has direct knowledge of, and relation to, its environment as a result of natural laws (Duchon, Kaelbling, & Warren, 1998). I will note here that there is overlap with another school of psychology also referred to as Environmental Psychology, but this school oftentimes goes by the name *Environmental Psychology* (see Barker, 1968). Both schools emphasize "real world" studies of behavior as opposed to the artificial environment of the laboratory. In his work "Ecological Psychology: Concepts and Methods for Studying the Environment of Human Behavior" (1968), Barker argued that human behavior was radically situated; that is, you couldn't make predictions about human behavior unless you know the situation or environment a person was in. While listening situations and natural environments are relevant to my proposed stimuli, please know that my subsequent references to Ecological Psychology refer to the "Gibsonian" school.

Gibson championed the idea that observers sample information from the outside visual world using an active perceptual *system* rather than passively receiving input through their senses. Perceptual systems pick up the information they are tuned to (akin to "sympathetic resonance") through direct perception; i.e., perception is unmediated by inference or memory. For Gibson, the world contained “invariant” information that was directly accessible to the perceptual systems animals. Elemental stimuli are specified by complex invariants of supposedly primitive features. “A perceptual system does not respond to stimuli (although a receptor does) but extracts invariants” (Gibson, 1979). Gibson also coined the term “affordances.” Affordances refers to the interactive possibilities of a particular object or environment. The observer is tuned to the *invariant* properties associated with the *affordances* of an object.

Another concept important to Ecological psychology is *perceptual constancy*. I will attempt to explain this with examples. The brightness, color, and size of an object are perceived as being constant despite changes in the ambient light or the object’s location as it moves away or towards us. Similarly, we can identify a sound source (such as a familiar voice) independent of changes in the listening environment that influences the sound’s duration, intensity, or spectral makeup. We perceive the sound as constant, although the physical variables change as the object moves relative to the listener. Given an infinite possibility of different environmental conditions and, consequently, an infinite number of sensory patterns that stimulate the senses, we are able to perceive a given object with more-or-less constant properties.

Traditional Psychoacoustics and Ecological Psychology

When it comes to studying hearing disorders, there may be reason to believe we’ve gone “full circle” in terms of the test stimuli used. Prior to the advent of the electron tube, sounds were generated mechanically via sirens, pitch pipes, tuning forks, etc. Electronic technology has since allowed the researcher to create sounds with precise control of the physical variables (frequency, duration, and intensity) that are deemed important to the research question being answered. But despite our current understanding of the auditory periphery, there’s a lot we don’t about *listening*. If the physical variables alone accounted for all of hearing perception, there wouldn’t be reason to adapt the complex sounds we hear in everyday listening as test stimuli other than, perhaps, to demonstrate *external* validity.

Traditional approaches to studying hearing generally focused on the physics of the signal and physiological response on the peripheral auditory system. Stimuli typically consist of short sequences of pure or complex tones, noise bursts, or simple combinations of these sounds. Many psychoacousticians concentrated their efforts on understanding the specific manner in which acoustical energy is transduced into an electrochemical signal in the nervous system. Central issues have included thresholds of detection and discrimination, how frequency and intensity are coded, and how sound sources are localized. The advantage of traditional approaches is that one can gain an understanding of how peripheral auditory system functions. The disadvantage of this approach is that the stimuli and listening conditions employed (headphones, sound-attenuating booths, and anechoic chambers) are often unrealistic, leaving the larger question of environmentally and ecologically important cognition, listening behaviors, and the linkage between perception and action relatively unexplored. Many psychoacousticians were keenly aware that cognition influenced behavior and action in response to an acoustic signal. But rather than studying "confounding" phenomenon or investigating its relationship with sensation and psychology, the approach was to develop experimental controls or use statistical methods (such as signal detection theory) to factor out decision criteria or other consequences of cognition.

In any type of experimental setting, the validity of the findings is crucial to interpreting the results and drawing reasonable conclusions about the data. There are various types of validity, most of which have been extensively outlined by Campbell and Stanley (1963). The focus of traditional psychoacoustic research has been on internal validity, assuring that the experimental manipulations or independent variables are truly responsible for any changes in the dependent variable. Typically, researchers have gone to great lengths to assure that the experiments are conducted in a quiet, controlled environment with tightly-controlled stimuli. All of these precautions are requisites from the standpoint of maintaining the internal validity of the research.

External validity is the degree to which the results of an experiment would apply in other settings with other populations. For example, how might an auditory effect discovered by conducting an experiment in a sound booth hold up in a concert hall or a living room? The experimental control afforded by an anechoic room (for example) increases internal validity at the cost of external validity. If the critical question is to make inferences about the physiology of the peripheral auditory system, these questions become less of a concern. However, if one is interested in listening behavior, how listeners perceive and use acoustic information to guide action, then these questions become crucial.

If our perceptual abilities have evolved specifically to deal with the stimuli that occur in a natural environment, perhaps it is not surprising that there are differences in processing naturally occurring stimuli and those that are more artificial. In real-life conditions, perceptual information is not the sum of simple perceptions. For example, how an observer recognizes complex visual objects cannot be derived from research on the perception of simple geometric figures. Similarly, how a listener perceives complex acoustical *events* cannot be derived from research based on puretones, sounds built on a few simple tones, noise bursts, or simple combinations of these sounds. Gaver (1993) stated that the largest part of hearing research has been devoted to the study of *musical listening*; i.e., the study of perception of tonal sounds. In contrast to typical laboratory stimuli, *everyday sounds* are inharmonic, noisy, non-stationary, and have irregular temporal patterns. Because of their complexity, everyday sounds are difficult to describe in terms of frequency, amplitude, phase, and duration. Understanding the physics of acoustic events is not sufficient for understanding how we perceive these events. If one parameter in a physical model changes, many attributes can change as a result. In accordance with ecological psychology, elemental stimuli are specified by complex invariants of supposedly primitive features. It is the complex patterns of *change* that contain the information specifying an event perceptually.

It has been proposed (Gaver, 1993) that we study the perceptual dimensions of the *sources* as well as those of the sounds they make. From this perspective, we should treat complex acoustic variables as elementary. The study of perception should be aimed at uncovering the relevant dimensions of perception and the invariant perceptual information derived from them. What Gaver (1993) refers to as “everyday sounds” are much more complex and are rich in redundant and complementary information. Examples on the perception of everyday sounds are studies on the perception of bouncing and breaking sounds (Warren & Verbrugge, 1984) and on the sound of walking (Li, Logan & Pastore, 1991). Li, Logan and Pastore (1991) separated sound into three kinds of phenomena: 1) auditory events, 2) acoustic structure, and 3) perception. They stated that only by analyzing the relation among these phenomena can we understand the perception of everyday sounds.

Background: Cochlear Implants

A cochlear implant is a surgically implanted electronic device (prosthesis) that provides the sensation of hearing to a person who is profoundly deaf or severely hard of hearing. Cochlear implants directly stimulate the auditory nerve using electrical pulses. Ideally, this would allow the brain to perceive sounds as it would if the receptor hair cells along the basilar membrane were functioning. Two (of several) key components of a cochlear implant are the speech processor and a surgically-implanted array of electrodes. The electrodes are arranged in a "tonotopic organization." This organization is akin to the topography of the inner ear's receptor cells. Tonotopic organization (also referred to as frequency-to-place mapping) is how the inner ear codes for frequency; however, it should be pointed out that place mapping is but one way the ear-brain codes the perception of pitch (for example, see Warren, 1999 or Moore, 2007). When an electrical current is routed to an electrode, an electrical field is generated and the auditory nerve fibers proximal to that particular electrode are stimulated. The electric fields are not well-defined and this consequential spread of energy (or smearing) imposes limits on the number of electrodes that can be active (or electrically powered on) at any given instant. The speech processor determines which electrode(s) receive an electrical pulse as well as the amplitude and width of the stimulating pulses.

Modern cochlear implants have 12 – 24 electrodes or channels (per ear) to replace the approximately 3500 inner hair cells that innervate the auditory nerve. Despite this and other limitations, the sound quality delivered by a cochlear implant is often good enough that many users can understand speech in quiet conditions. In noisy conditions, speech understanding often remains poor. At worst, the cochlear implant allows many people to recognize warning signals or have awareness of sounds in the environment. Hearing through a cochlear implant is far from natural hearing and (to date) offers only very limited appreciation of musical melody.

For the cochlear implant user, listening ability is compromised because of certain limitations (two of which are explained above) and, furthermore, because the recipient's auditory nerve itself may be damaged. Traditional stimuli used to investigate hearing lack complexity and, consequently, provide *impoverished* input to the auditory periphery. The cochlear implant recipient receives impoverished information (at least at the neural level) regardless of stimuli complexity. But in our quest to improve cochlear implant efficacy, researchers continue to focus on pitch and loudness discrimination, localization, etc. by use of simple stimuli. Without doubt, many "traditional" studies have provided much useful information, but the assumption is that the elemental information for sound is contained in the readily-measured and easily-manipulated physical variables (frequency, intensity, and duration). Can we truly advance cochlear implant technology by continuing to test with "impoverished" stimuli or by using stimuli that don't replicate our everyday, natural environments?

Proposed Stimuli

This section of the paper addresses the stimuli that I propose for performing speech testing in noise. Features of the proposed stimuli used in my Veridical Reality System (VRS) that encompasses "loud speech" (to be explained), video, and what I refer to as *informative* sentences. The same techniques could easily be applied to non-speech sounds built on the "elemental" acoustic variables suggested by Gaver (1993). Although these non-speech sounds could provide valuable insight into hearing with cochlear implants, I have limited this discussion to speech sounds and speech comprehension in noise.

Nowadays it is both economically and technically feasible to record and reproduce *everyday sounds* with exceptional fidelity. But "high-fidelity" alone only addresses the physical variables of sound that have been tightly controlled in traditional psychoacoustic experiments. While faithful playback would be requisite to both internal and external validity, the *information* a stimulus carries (and how?) aren't necessarily captured by the recording chain. It is for this reason that I have devised methods for recording acoustical *events* that can then be used in a controlled, laboratory setting. Some of my suggestions come from ideas that may seem so obvious as to be naïve; whereas, at least one of my recording techniques is novel enough that I didn't find reference to it in a literary review. By careful integration of scientifically-based research, ideas derived from

Ecological Psychology, and a newly proposed voice-recording technique (to be described), stimuli can be developed that represent “real world” listening. The stimuli that I propose should transcend the limitations of current speech tests, particularly when testing cochlear implant patients in noisy environments that are representative of real-world scenarios.

While it is not possible to record or replicate the unlimited number of all real-world scenarios, it is feasible to record and replicate (acoustically) several listening environments that many cochlear implant users are likely to encounter, and then draw general conclusions based on the results of listening tests performed in these life-like environments. Although capturing a sound’s physical attributes would seem trivial given today’s technology, designing and implementing a controlled but “real-world” environment in the laboratory is not a trivial task. If we accept that sound-source location affords meaningful information about an auditory event, then accurate replication of real-world environments should include not only faithful reproduction of a sound’s frequency, intensity, and duration, but also include each sound’s respective location in three-dimensional space. Improvements toward aural realism have been made possible by use multi-channel surround systems (to include the VRS). A commercially available, 8-channel surround system known as “R-Space” (www.revitronix.com) and accompanying live recordings were developed for use in auditory research environments.

Much of what I know regarding the R-Space™ comes from personal communication with R-Space’s co-developer Lawrence (Larry) Revit. It is my understanding that the original intent of the R-Space system was to study and demonstrate the purported advantages of different hearing aid microphone configurations (e.g., directional versus omnidirectional) and noise-reduction strategies in an exemplary “real-world” environment. I had interest in the R-Space, not only because of its application to hearing research, but because years earlier I had independently designed a similar system for presenting stimuli in a continuous surround of background noise but with a slightly different application in mind. One study that made use of my prototype surround system was ultimately published in Noise & Health (Carmichel, Harris, Story, 2007).

The R-Space™ was designed to fit within the confines of typical sound-attenuating booths. (While this provides some control of the acoustical environment and fits within the space of many hearing test booths, the R-Space's small, 2-foot radius does impose constraints.) The R-Space system presents background noise made from 8-channel on-site (live) recording. While considerable thought was put into recording and the playback of a real-world environment (Lou Malnati's Restaurant in Chicago, IL), much less emphasis has been placed on the material used as the target, or signal, stimuli. This isn't to say that the variety of speech stimuli currently being used weren't recorded in a controlled environment, or that they weren't balanced in terms of spectral content or ease of comprehension; they were. My caveat is that the speech material focuses more on *internal* validity than external validity.

What I propose is the development of stimuli that includes the subtle information that could be relevant to *everyday* listening. Capturing subtle information is not a matter higher fidelity recordings of human speech (hi-fi is the easy part with today's technology), but is a matter of capturing aspects of *language* that better represent speech production in noisy environments. While it may seem simple and straightforward to record human voice, aspects of "normal" speech in noise have not always been taken into account. Speech stimuli are typically recorded in a background of quiet and at the talker's normal speech-production level. Making speech louder is trivially accomplished by increasing the power output to the headphones or loudspeaker used to present the signal. If the experimenter wishes to increase the level of signal so that it is above the background noise an additional x dB, all he/she has to do is increase amplifier gain an additional x dB. But this method of increasing speech (signal) level assumes that improvement in speech comprehension is solely a function of SNR. This approach ignores the possibility that the *way* in which a person talks above noise may provide information important for speech understanding. Because I am also proposing the inclusion of lipreading as a part of my stimuli (this is nothing new, but it is one way of getting *meaningful* speech recognition scores in a background of moderately loud restaurant noise), the possibility of *loudness* perception based on a speaker's lip movements should be taken into account. In some instances, the detection of loudness variations may not change in concert with sound *intensity* because of the listener's hearing impairment, automatic gain control circuitry within a listener's hearing device, or background noise level.

In the discussion that follows, the *talker* is the person whose voice will be used to create the speech stimuli. The background noise, independent of the talker, can be multi-track recordings that simulate any number of environments. The VRS uses a combination of live recordings and *auralization*. While the techniques used for recording and presenting background noise are not the focus of this paper, they will be discussed in a forthcoming paper. The focus here is on generating the target (signal) stimuli. The proposed stimuli can be used with existing surround systems (e.g., the R-Space™ being used at ASU and elsewhere or with my Veridical Reality System).

It shouldn't take a lot to convince anyone that we tend to raise the level of our own voice in a noisy environment; this phenomenon is known as the Lombard effect (see, for example, Lane & Tranel, 1971). Exactly how much louder as a function of background noise level is a question to be asked if we are to create realistically loud speech. Tufts & Frank (2003) showed that talkers' voice levels increase, on average, 5 dB for every 10 dB increase in background noise level. For my research, this data is potentially useful. However, these data were obtained while the talkers were donning hearing protection or talking in a background of white noise. See Figure 1 for a comparison of white noise, pink noise, and restaurant noise. It should also be note that monitoring one's own voice under hearing protectors isn't the same as talking in a background of noise. A convincing demonstration of this is to simply cover both ears with your hands: Your voice will sound muffled and unnatural, thus precluding "naturalistic" monitoring of your own voice.

This may seem like a lot of fuss over speech "loudness" when all we have to do is adjust the gain of an amplifier powering a loudspeaker to adjust the signal's level. But I argue that the *way* in which we speak in noise conveys information, and subtle features of the talker's voice and face should be included in order to make listening in background noise more realistic. In general, when speech is produced in a background of noise, acoustic and prosodic features (overall level, fundamental frequency, formant frequencies, and spectral composition) of the speech change relative to speech produced at normal conversation levels in quiet (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). The trick, however, is how to record “loud speech” without recording the background noise that elicits this *natural* increase in speech loudness. Note that I’m using loudness in lieu of intensity: Without direct measurement, I cannot state if the perceived loudness is due solely to an increase in SPL.

There are ways of getting a talker to speak louder. The obvious would be to 1) ask the talker to speak louder (but then, how loud?), or 2) place the talker in a background of noise. Measuring an increase in the talker’s level as a function noise level presents something of a challenge because the change in the talker’s voice level may not be discernable from the increase in noise level. In other words, the talker’s own voice contributes minimally to the combined signal-plus-noise level. Consequently, we can’t accurately determine an increase in signal level from the noise level. For example, a 65 dB SPL signal (the talker) added to 70 dB SPL of uncorrelated background noise results in a combined level equal to 71 dB SPL. Discerning the signal level from the overall level is even more difficult when the signal and noise are both fluctuating and have similar spectral characteristics (as is the case with speech in speech noise).

I have developed a recording technique that makes it possible to record naturally loud speech in a background of noise but without recording the noise. Separating the signal from the noise will, then, allow the researcher to use the “naturally loud” speech signals (whether they are words or sentences) in a variety of ways and background conditions. The question is how to effectively separate the signal from the noise: A highly direction microphone will not work, especially if there’s noise behind the talker; and signal processing and noise-reduction techniques don’t work well in dynamic environments, particularly if the background noise is conversation. Adding noise through insert phones or headphones doesn’t allow the talker to self-monitor their own speech; that is, they don’t sound “natural” to themselves because of occluded ears. However, if a talker *could* self-monitor his/her own voice so that it truly sounds natural, then the application of noise via insert phones may very well work.

In order to elicit natural speech, the talkers must hear themselves as though there were no insert phones. Insert phones change the way we hear our own voice for two reasons: 1) occluding the outer ear canal and concha destroys a naturally-occurring peak in hearing sensitivity at around 3000 Hz, and 2) low-frequency sounds from our own voice are enhanced because the reduction of ear canal volume [referring to physical volume in units cubed, not intensity]. I have measured these sensitivity and pressure changes caused by insert phones placed in my own ears. Swept frequency measurements and Maximum Length Sequencing (MLS) techniques were used to measure the response peak due to ear canal resonance. One swept-frequency measurement (obtained from my ear) is shown in Figure 2. By use of a specially developed microphone, I am able to produce vowel sounds and compare the pressure differences of an occluded versus open ear. The reason for using one's own voice instead of using an external generator is this: The sound pressure that gets amplified via occlusion comes from the talker's own voice; the sound is transmitted to the ear via bone conduction.

Frequency response aberrations caused by occluding the ear canal can be corrected for by use of a parametric equalizer. Real-time processing is requisite so that talkers can self-monitor their own voice as they would normally. After filtering (equalization) the talker's own voice will be heard without frequency response anomalies; that is, the talkers hear their own voice as though both ears were open. Background noises are independently controlled in level. Talkers' own voices can only be made louder to themselves by increasing their voice level; i.e., as they would in noise (Note: changing the *level* of the background noise does not affect the fixed *gain* of the self-monitoring system). Again, gain is fixed (unity or 0 dB) for voice self-monitoring, but the background noise level can simultaneously and arbitrarily be adjusted. For initial tests, I used stereophonic and surround sound (HRTF-processed) recordings of restaurant noise, presented through the insert phones, as the background noise. Clearly, other environmental noises could be used, and their effects on a talker's voice production could be measured. Using the method described above, talkers hear their own voice and background noises "naturally," whereas the recording microphone, placed 0.7 meter away (my setup's signal-to-listener distance), only records the talker's voice. A talker's voice, as it would be produced in noise, is then captured (minus the noise) as the "new" stimuli. Video recordings of the talker are used to capture visual information that could be relevant to our ability to comprehend speech in noise. Subtle changes in visual and auditory cues as they change with noise level can be quantified and observed.

While there is nothing new or novel about incorporating video, I am unaware of any investigations using cochlear implant patients' in a surround of *uncorrelated* background noise combined with video presentation of the talker. One could also study the effects of *simulated* cochlear implant hearing with visual cues in a natural environment. It has been known for some time that lipreading is useful for comprehending speech presented in a background of noise. For example, Sumby & Pollack (1954) showed that visual cues could aid speech comprehension to the same degree as a 15 dB improvement in SNR. Sounds with acoustic features that are easily masked by white noise (for example, the voiceless consonants /k/ and /p/) are easy to discriminate *visually*. To quote Miller & Nicely (1955), "The place of articulation, which was hardest to hear in out test, is the easiest to see on a talker's lips. The other features are hard to see, but are easy to hear." Most prior tests, however, used white or pink noise as the background noise, and the noise originated from a single loudspeaker or earphone. A comparison of white, pink, and restaurant noise is shown in Figure 1. As I have stated, I am interested in creating stimuli that includes "naturally loud" speech in natural settings (e.g., speech production in a noisy restaurant). But the key is to isolate the signal (talker) in such a way that the speech and video, minus the background noise, can be presented as a stand-alone stimulus.

There is a plethora of literature surrounding the benefits of lipreading. In addition to movement of the lips as an aid to recognizing specific phonemes, vision can affect the perception of a co-articulated speech stream. Green & Miller (1986) showed that just by seeing a person talk quickly or slowly was sufficient to change the judgment of a series of "pea" being repeatedly spoken as "pea" or "bee." Capturing naturally spoken language in a natural environment may, however, require more than just video-taping the talker. To date, I am unaware of studies using uncorrelated surround noise plus video to study cochlear implant efficacy. It is entirely possible that visual cues can affect more than just speech comprehension. A study showing the reduction of stress when a listener is aided by lipreading could be interesting: It is possible that visual cues, regardless of speech comprehension advantages, could reduce listener stress in a difficult listening environment. Another motivation for including lipreading at this time is this: There is ongoing research using a surround of uncorrelated background noise via the R-Space.

However, the realism of the background noise is tainted (the natural level has been reduced to 60 dBA) because listening in the noise at real-world levels (say 70 dBA) oftentimes makes it *too* difficult for cochlear implant patients to comprehend arbitrary sentences. Reducing the background noise to 60 dBA or increasing the signal level to 80 dBA yields meaningful test scores with a +10 SNR (using our current test protocols and materials) but at the expense of *external* validity. Conversely, other studies (reference) presented the background noise at 70 dBA, but then elevated the signal level to maintain a +10 or greater SNR. "Normal" speech presented at 80+ dBA is *not* realistic. The psychoacoustician might call the addition of visual cues a confounding variable, whereas the ecological psychologist might discover synergistic affects of available information in the environment when testing is performed in life-like simulations of the real world that include both video and audio.

Conclusion

Capturing events that represent real-world events and likely scenarios can be reproduced with certain realism in the laboratory. Capturing subtleties, such as talker voice level as a function of background noise level, could make video and audio stimuli more realistic. Although we might not be sensitive to these subtleties when sufficient information is available to us, hearing-impaired individuals might make use of visual cues in ways that have not been explored. At the very least, recording techniques make it possible to capture real-world sound events that can be played back time-and-time again. Systematic reduction or elimination of available information can be accomplished when the stimulus contains all of the essential information in the environment. What is deemed *essential* in the stimuli is left to the researcher; at least highly realistic simulations of our environment will provide this information, whereas an artificial stimulus may or may not provide subtle information, regardless of its complexity. My goal is to produce life-like stimuli that may prove valuable to future hearing research.

Additional Notes

More than a few studies have used normal-hearing listeners and vocoded speech to simulate listening with a cochlear implant. Distorting both the signal and surround background noise via vocoding could allow for realistic cochlear implant simulations when using normal-hearing subjects in noise. A potential study would be to vocode, in real-time, the background noise and speech signal (with or without video), and then to compare the results to studies using vocoded speech presented in a background of monaural pink noise. To date, I am unaware of any studies that have vocoded naturally occurring sounds (such as those suggested by Gaver, 1993) or the uncorrelated surround noise (such as restaurant noise) to simulate listening with a cochlear implant.

Another issue that applies to speech stimuli and natural environments is *what* is being said. For example, compare the phrase “The train leaves in fifteen minutes” to “The plane leaves in fifty minutes.” Both sentences contain potentially valuable information, and it would certainly be important for the listener to distinguish between “fifty” and “fifteen.” Subtle information in the environment (beyond lipreading) may provide an aid to world discrimination while in difficult listening situations. I tend to view speech stimuli that don’t have contextual meaning with a dubious eye. As an example, would the nonsense sentence “He had the saxophone of a quilted sugar razor” be encountered in an airport terminal or train depot (or encountered at all!)? [And I’ve heard worse…]

The experimental design and statistical analysis using speech signals with video in a surround of naturally-occurring background noise should also be considered. When video recording is involved, the tendency might be to use a t-test to compare speech comprehension with and without the visual cues. If we were to compare makes or models of cochlear implants, all models would receive equal benefit from the video, and the testing would be performed in a like-like environment with typical (not reduced) noise levels. Other considerations include the development of *natural* stimuli that work in concert with adaptive test protocols, forced-choice methods of testing, ergonomic subject interface devices, and automated data presentation and data collection systems. My approaches to all of these will be topics of future ramblings and will be posted on my forthcoming website, www.cochlearconcepts.com.

Figures

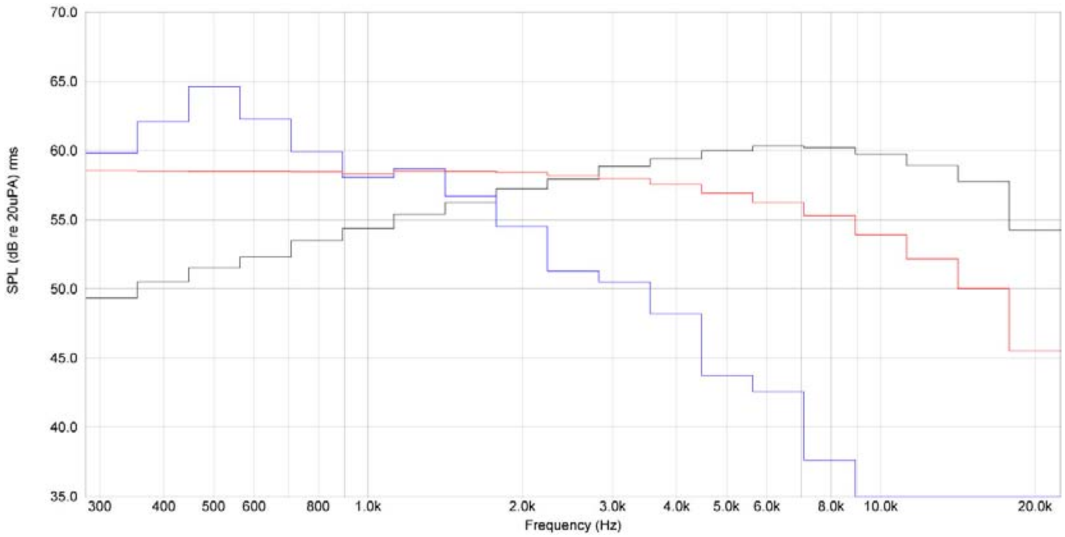

Figure 1. Comparison of 70 dBC white noise (black line), 70 dBC pink noise (red line), and 70 dBC restaurant noise (blue line) using third-octave analysis. Pink noise has equal energy per octave band, not uniform amplitude across frequencies (white noise has uniform amplitude across frequencies). Effects of masking would be different for these three noises even though they are all at 70 dBC.

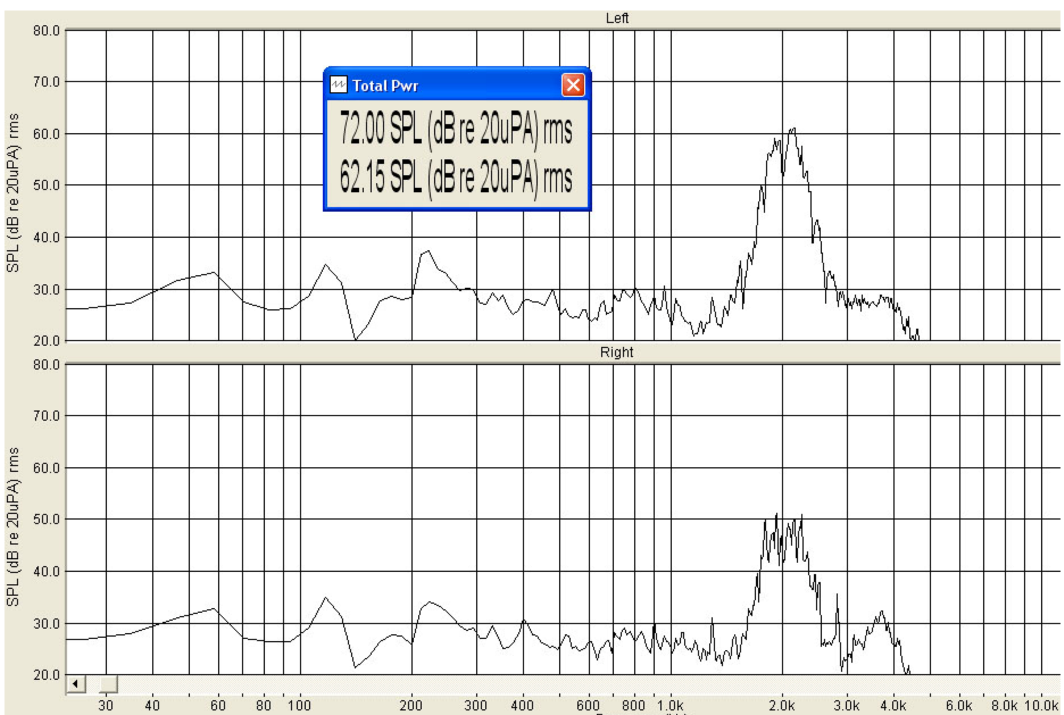

Figure 2. Peak in response due to outer ear canal resonance. This peak is lost when donning insert phones or earplugs, thus affecting the way we hear our own voice and other sounds with open ears.

References

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.

Campbell, D. T., and Stanley, J. C. (1963). *Experimental and quasi-experimental design for research*. Boston: Houghton Mifflin.

Carmichel, E. L., Harris, F. P., and Story, B. H. (2007). Effects of Binaural Electronic Hearing Protectors on Localization and Response Time to Sounds in the Horizontal Plane. *Noise & Health*, 9(37), 83-95.

Duchon, A. P., Kaelbling, L. P., and Warren, W. H. (1998). Ecological robotics. *Adaptive behavior*, 6(3/4), 473-507.

Gaver, W. W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5(1), 1-29.

Gibson, James J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, James J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Green, K. P., and Miller, J. L. (1986). On the role of visual rate information in phonetic perception. *Perception & Psychophysics*, 38, 269-276.

Lane, H., and Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *J. Speech Hear. Res.*, 14, 677-709.

Li, X., Logan, R. J., and Pastor, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *J. Acoust. Soc. Am.*, 90, 3036-3049.

Miller, G. A., and Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.*, 27, 338-352.

Moore, B. C. J. (2007). *An introduction to the psychology of hearing*. Elsevier Ltd.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *J. Acoust. Soc. Am.*, 84, 917-928.

Tufts, J. B., and Frank, T. (2003). Speech production in noise with and without hearing protection. *J. Acoust. Soc. Am.*, 114(2), 1069-1080.

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge, UK: Cambridge University Press.

Warren, W. H., Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 704-712.