DIGITAL SOUND SYNTHESIS, ACOUSTICS, AND PERCEPTION: A RICH INTERSECTION

John M. Chowning

The Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University
jmc@ispchannel.com

ABSTRACT

The early years of digital sound synthesis were filled with promise following Max Mathews' publication in 1963 of his pioneering work at Bell Telephone Laboratories [1]. The digital control of loudspeakers allowed for the production of any conceivable sound *given the correct sequence of numbers (samples)*. Producing the correct sequence of numbers, however, turned out to be a formidable task. Acoustics and psychoacoustics, the first a well-developed field of knowledge and the second less so, did not provide information at the level of detail required to simulate even the simplest sound of an acoustic instrument.

The enormous potential of digital synthesis counterpoised with an enormous knowledge deficit were the initial conditions for interdisciplinary research that continues to this day. Discoveries have been made and insights gained that are of consequence in the general field of digital audio.

1. INTRODUCTION

Perceptual studies in regard to sound follow a long tradition most often associated with the hearing sciences in medicine, but occasionally in engineering sciences, most notably at Bell Telephone Laboratories (BTL), and independent laboratories at a number of locations. With the early association of computer sound synthesis and music in the early 1960s, the study of workings of the auditory system became absolutely essential for two reasons:

- the pragmatic need to understand how the ear responds in order to make efficient use of extremely expensive computer cycles and memory, and
- the functional need to decipher the attributes of complex sound for the purpose of sound design and artificial acoustic space design.

This paper describes one aspect of the author's own work having to do with loudness and how it related to and grew out of perceptual/acoustic studies performed at BTL.

2. THE EARLY YEARS

2.1. The Importance of Psychoacoustics

Max Mathews, generally acknowledged to be the "father" of computer music, wrote in 1963, "There are no theoretical limitations to the performance of the computer as a source of musical sounds, in contrast to the performance of ordinary instruments [1]." For anyone who had generated sound and tried to simulate acoustic instrument sounds using the analog technology of the time, this was a bold claim (that found its substantiation in sampling theory). It was also a bold invitation to explore the potential of the rapidly evolving digital technology. He continued, "…the range of computer music is limited principally by cost and by our knowledge of psychoacoustics. These limits are rapidly receding." It turned out that the reduction in cost of computing was a far more tractable problem than that of increasing the knowledge about the auditory system. While there has been a vast amount of research in the field of psychoacoustics, only a small amount has had direct relevance to the art of generating sound and locating sound in simulated acoustic spaces.

2.2. Synthesis from Analysis

Jean-Claude Risset's training in both physics and music performance/composition made him the ideal candidate to work with Mathews in extending acoustic/psychoacoustic theory beginning in 1964 at Bell Telephone Laboratories. They chose to use the computer to analyze real instrument tones and then synthesize those tones using the physical description derived form the analysis [2]. The synthesized tone was then compared to the original as a way of measuring the success of the analysis and synthesis.

This approach produced remarkable results in that many of the synthetic tones were indistinguishable from the original. In the process, Risset made some insightful observations regarding the evolution of instruments' spectra through the course of tone. Of particular interest was his study of trumpet tones. The timbral "signature" of brass tones was elusive. Standard acoustics studies suggested recipes which when rendered with the precision of the computer produced tones that had no resemblance to the intended tone. Unlike the clarinet, where the

odd harmonic emphasis serves as a "signature" for the family, there is no similar strongly unique harmonic emphasis for the trumpet. In fact, its so-called steady-state can easily be confused with that of an oboe. Risset discovered the "signature" of the trumpet in the attack portion of a tone.

The evolution of the harmonic amplitudes of the spectrum during the attack portion is rapid, complicated, but patterned. As the overall amplitude of the tone the grater the relative contribution of the higher harmonics (see Fig. 1).

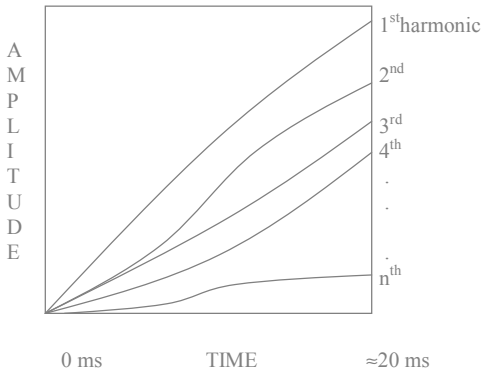

Figure 1. *Evolution of harmonics for a trumpet tone.*

The pattern, however, is not discernable by the "ear" as it is heard as a totality according to the Gestalt "law of common fate" where components moving in the same direction are perceived as a group. Risset then vastly reduced the data required for effective simulation by formulating an algorithm that preserved the indispensable attribute of the trumpet's signature: *the centroid of the spectrum shifts up in frequency with the increase of overall amplitude of the tone.*

Considering all of the physical complexity in the production of a simple trumpet tone (now well understood as a result of powerful physical models), Risset's isolation of the most important perceptually relevant feature was a milestone. His work demonstrated the selectivity of the auditory system for certain, but not all, of the physical complexity within a tone in order to identify the source as *a trumpet*.

It must be pointed out that understanding the perceptual relevance of the various physical acoustic properties of a tone was not only of importance to the field of psychoacoustics, but it was essential to the nascent area of music composition/sound synthesis. Without this work, computer music would have been vastly reduced in quantity and quality because of the cost of computation and the complexity of Fourier synthesis and/or digital signal processing.

2.3. Analysis by Synthesis

The breakthrough insight in the development of Frequency Modulation Synthesis (FM) by this author was directly related to the work of Risset and Mathews sited in the above studies. The means by which complex time-varying tones can be generated by this technique was discovered in 1966 and described in 1973 [3].

It was realized early on following the discovery, that the inherent spectral properties of the process were closely aligned to perceptually relevant attributes of a number of acoustic instruments. This was determined by subjective evaluations and therefore became known as *analysis by synthesis* since the manipulation of FM parameters (with ever-increasing insight) could produce credible simulations of a variety of drums, gongs, and woodwinds, for example, (but *not* the brass family). A successful synthesis implied the presence of properties or characteristics that were found in physical analyses of the sort performed by Mathews and Risset.

Broad generalizations were made based upon this research, one of which became especially relevant to the later discussion regarding loudness and auditory perspective. In blown and bowed instruments the number and prominence of partials (for the most part harmonic) increase and decrease proportionately with overall amplitude, and in freely vibrating spectra (plucked, struck and for the most part inharmonic) the number of partials decrease in number as the overall amplitude of the tone decreases.

As shown in Fig. 2., FM synthesis can produce time varying spectra when the modulation index is allowed to vary as a function of time. In fact, when the envelope shape that controls the overall amplitude of the tone is used to control the modulation index through time, a convincing brass-like tone can be produced (with appropriate scaling and parameter sitting). This essential insight that resulted from Risset's trumpet studies is what moved the FM technique from one of local interest to one of widespread use.

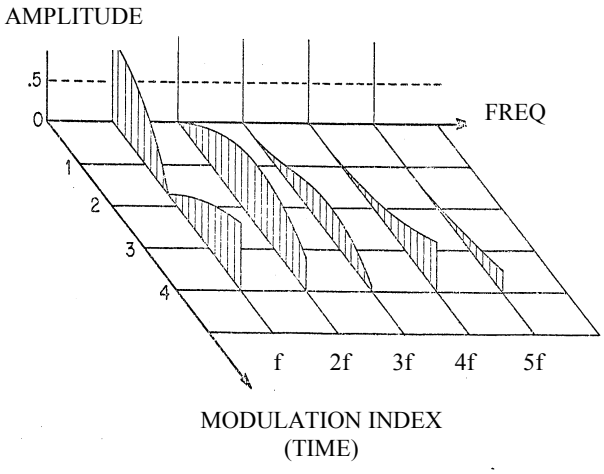

Figure 2. *Evolution of partials as the modulation index changes in time (original hand-drawn Figure from 1972).*

3. AUDITORY PERSPECTIVE

The perception of sound in space remains an important issue in sound emanating from loudspeakers, whether prerecorded, from digital instruments, or from computers. In the simplest case a listener localizes the emanating sound from points defined by the position of the loudspeakers. In all other acoustic settings the listener associates a sound source with horizontal and vertical direction and a distance. The auditory system seems to map its perceived information to the higher cognitive levels in ways

analogous to the visual system. Acoustic images of great breadth reduce to a point source at great distances, as one would experience listening to an orchestra first at a distance of 20m and then at 300m, equivalent to converging lines and the vanishing point. Sounds lose intensity with distance just as objects diminish in size. Timbral definition diminishes with distance of a sound from a listener just as there is a color gradient over large distance in vision. Therefore *perspective* is as much a part of the auditory system as it is of the visual system. It is not surprising that the two systems should have evolved in a way that avoids conflict of sensory mode in comprehending the external world since many visually perceived objects can also be sound sources. The auditory location of sources can be especially important to survival, for example the proximity of a mother voice or the

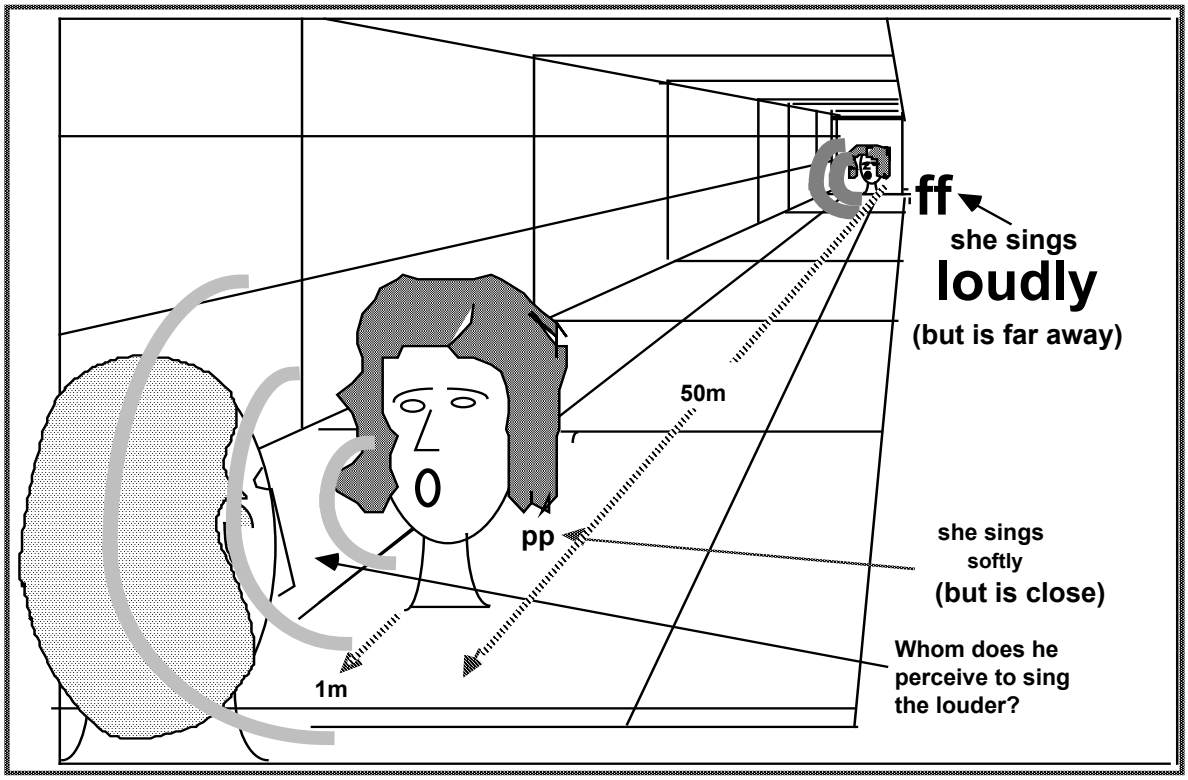

Figure 3. *As the visual system determines the distant singer's real size based upon perspective, the auditory system must determine loudness by an analogous strategy. The sound from the close singer arrives at the listener's ear having the greater physical intensity, yet the listener hears the distant singer as the louder.*

growl of a lion at a distance or close at hand, or the approach of a fast moving automobile. While not nearly equal in precision, the auditory localization system has two attributes that the visual system does not have, a 360 degree scan and it functions in darkness.

Auditory perspective, the perceived position of sound in space, is composed of important acoustic and psychoacoustic dimensions.

3.1. LOUDNESS

As noted above, the spectrum of an instrument tone changes as the overall amplitude or loudness of the tone changes. The amount and rate of change depends upon the force applied by the performer (bow or breath pressure, strike force, etc.). Life experience confirms that this change in the spectrum is easily perceived: a softly played tone at a given pitch and duration is different in *tone quality* as well as loudness when compared to a loudly played tone. However, there are common contexts where a difference in the overall amplitude of a tone can be perceived without a corresponding change in the spectrum. Listeners located at different distances from a tone source perceive a difference in overall amplitude but do not perceive a difference in *loudness*. Commonly thought to be the perceptual correlate of physical intensity [4], loudness is a more complicated percept involving more than one dimension. In order to reveal this we can imagine the following experiment.

A listener faces two singers, one at a distance of 1m and the other at a distance of 50m. The closer singer produces a **pp** tone followed by the distant singer who produces a **ff** tone. Otherwise the tones have the same pitch, the same timbre, and are of the same duration. The listener is asked which of the two tones is the louder (See Fig. 3)?

Before speculating about the answer, we should consider the effect of distance on intensity. Sound emanates from a source as a spherical pressure wave (we are ignoring small variances resulting from the fact that few sources are a point).

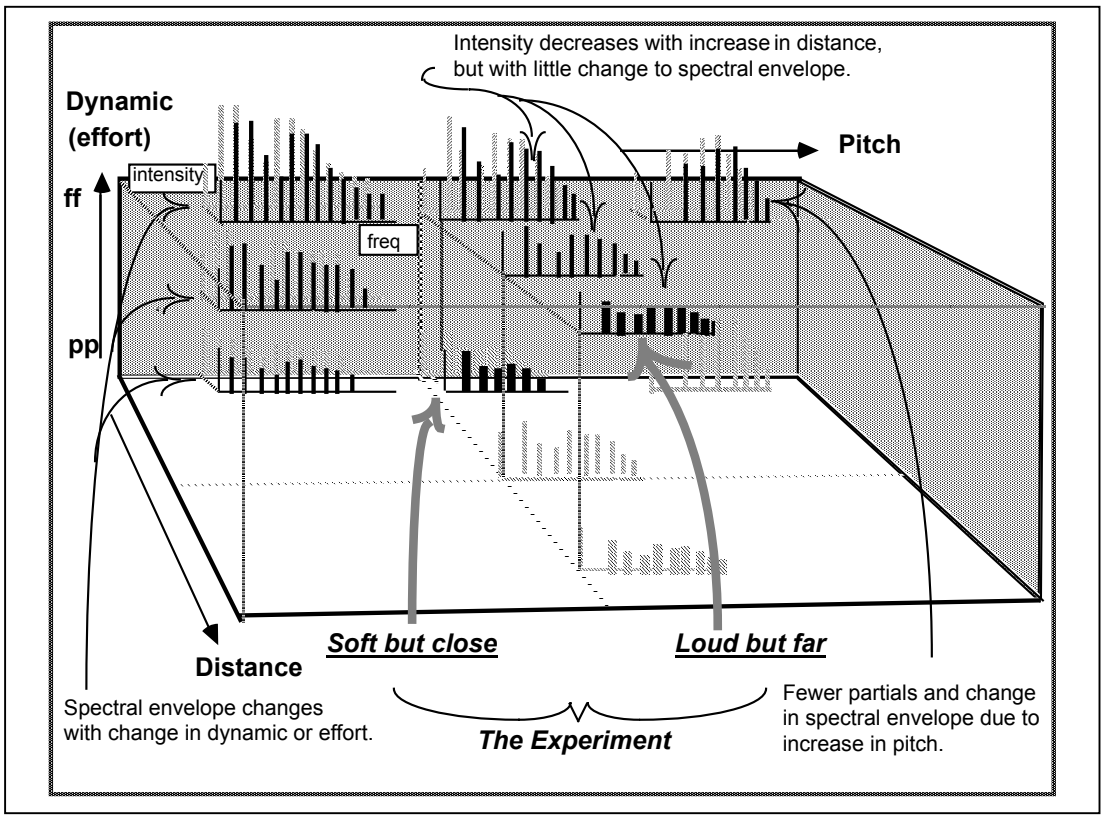

Figure 4. *Here we see the difference in overall intensity and spectral envelope between the tone that is soft and close and the tone that is loud but far, as projected in a five dimensional space.*

As the pressure wave travels away from the source the surface area of the wave increases with the square of the distance (as the area of a sphere increases with the square of the radius). The intensity at any point, then, decreases according to the inverse square law: $1/d^2$.

The distance in the experiment is 50m which will result in a decrease of intensity of $1/50^2$ or 1/2500 the intensity of the same **ff** tone sung at a distance of 1m. The listener, however, is asked to judge the relative loudness where the closer tone is a **pp** rather than **ff**. Let us suppose that the intensity of the **pp** is 1/128 that of the **ff**. The greater of the two intensities then is the closer **pp** and by a large amount. If loudness is indeed the perceptual correlate of intensity then the answer to the question is unambiguous. However, the listener's answer is that the second tone at 50m is the louder even though the intensity of the closer tone is about 20 times greater. How can this be so?

3.1.1. Spectral Cues

In the definition of the experiment it is stated that the timbre of the two tones is the same. The listener perceives the tones to be of the same timbral class: soprano tones that differ only in dynamic or vocal effort. In natural sources the spectral envelope shape can change significantly as pitch and energy applied to the source changes. In general, the number of partials in a spectrum decreases and the spectral envelope changes shape as pitch increases, that is the centroid of the spectrum shifts toward the fundamental. Similarly, the spectral envelope changes shape favoring the higher component frequencies as musical dynamic or effort increases, the centroid shifts away from the fundamental.

Fig. 4 represents a generalization of harmonic component intensity and spectral envelope change as a function of pitch, dynamic (effort), and distance. Because of the high dimensionality involved, a representation is presented where two-dimensional spaces (instantaneous spectra) are *nested* in an enclosing three dimensional space. The position of the origins of the two dimensional spaces are projected onto the 'walls' of the three dimensional space in order to see the relative values. Nesting spaces can allow visualization of dimensions greater in number than three, an otherwise unimaginable complexity*.

Now we can understand how the listener in the experiment was able to make a judgment regarding loudness that controverts the dominant effect of intensity on perceived loudness. Knowing the difference in timbral quality between a loudly or softly sung tone, reflecting vocal effort, the listener apparently chose spectral cue over intensity as primary. But what if the two tones in the experiment were produced by loudspeakers instead of singers and there were no spectral difference as a result of difference in effort? Again, the answer is most probably the distant tone even though its intensity is the lesser of the two - **if** there is reverberation produced as well.

3.1.2. Distance Cue and Reverberation

The direct signal is that part of the spherical wave that arrives uninterrupted, via a line of sight path, from a sound source to the listener's position. Reverberation is a collection of echoes, typically tens of thousands, reflecting from the various surfaces within a space arriving indirectly from the source to the listener's position. The intensity of the reverberant energy in relation to the intensity of the direct signal allows the listener to interpret a cue for distance. How does our listener in the experiment use reverberation to determine that the distant tone is the louder?

If, in a typical enclosed space, a source produces a sound at a constant dynamic, but at increasing distances from a stationary listener, approximately the same amount of reverberant energy will arrive at the listener's position while the direct signal will decrease in intensity according to the inverse square law. The source will be perceived by the stationary listener to have constant loudness as its distance increases from 1, 2, 3… etc. It is the constant intensity of the reverberant energy that provides the listener with the percept of a constant loudness at all distances even though the direct signal energy decreases as distance increases. The effect can be called **loudness constancy.** An analogous phenomenon, **size constancy,** occurs in the visual system. The perceived size of an object depends upon perspective and allows judgments to be made about size that do not necessarily correlate with size of the retinal image. In Fig. 5, we can see what is required to produce constant image size at the retina and constant intensity for the listener. The distant image **is** the same size as the closest.

Auditory perspective is not a metaphor in relation to visual perspective, but rather a phenomenon that seems to follow general laws of spatial perception. It is dependent upon loudness (subjective!) whose physical correlates we have seen to include spectral information and distance cue, in addition to intensity. Further, the perception of loudness can be affected by the 'chorus effect' and vibrato depth and rate in a very subtle but significant manner [5].

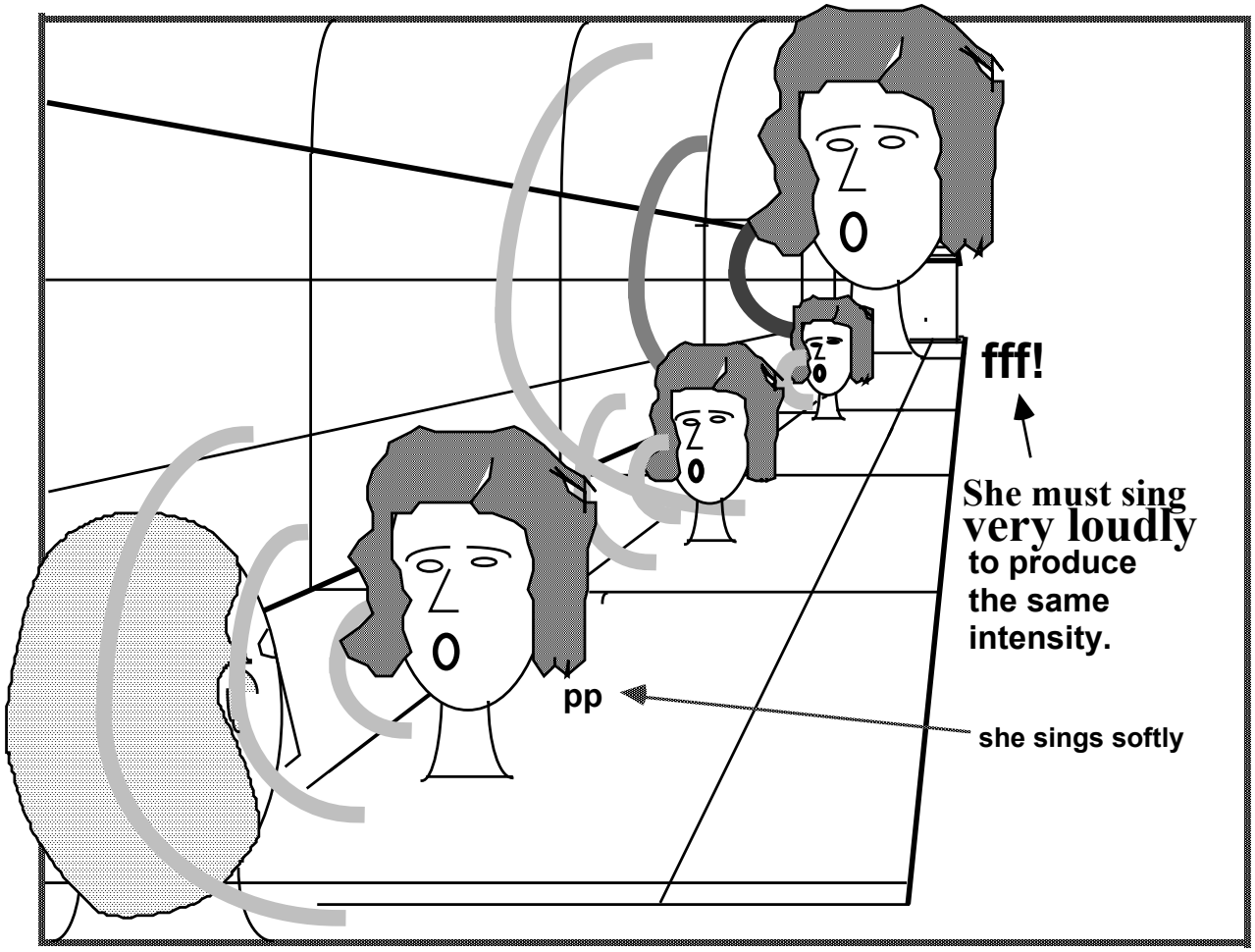

Figure 5. *The sound coming from a loudly sung tone from far away is the same physical intensity as that of the softly sung tone close by. But the spectral brightness gives an indication of the distant singer's loudness. If size constancy and loudness constancy were exactly analogous, the distant singer would look thus.*

The listener in the experiment, then, used all the information available, spectral cues, distance cues, and intensity, to make a determination of loudness at the source. When deprived of spectral cues the distance cue would provide the indication of the loudness at the source. Were there no reverberation present as in an anechoic chamber, then intensity would be the only cue as to loudness and the answer to the question would then be that the *closer* of the two is the louder.

Computers can, of course, be programmed to extend the dimensions of loudness beyond intensity thereby providing the listener with a percept of loudness vastly more subtle and **natural** than that provided by intensity alone. Given two loudspeakers (of even modest quality) on either side of a computer monitor, attention to details of sound projection can provide an auditory dimensionality unmatchable by current monitor technology. Today's computers and networks have sufficient power and bandwidth to achieve high quality sound projection and the perceptual importance of these dimensions of loudness can not be over-emphasized, yet their use is not widespread.

4. CONCLUSION

The study of acoustics and the perceptual system has had a productive outcome. Beginning with Risset's early acoustic/perceptual studies on trumpet tones and his critical insight into the nature of partial amplitudes and their relation to timbral authenticity, leading to and enriching FM synthesis, and then to the subsequent refinement of the definition of loudness to include the concept of auditory perspective, psychoacoustics has been a surprisingly fertile field in relation to computers and sound/music.

The domain of sounds to which these issues are relevant is not constrained to those similar to natural sounds, but may include all imaginable sounds. In fact, the understanding and exploration of these issues suggests somewhat magical musical/acoustic boundaries that cannot be a part of our traditional acoustic experience yet which can find expression through machines in ways that are consonant with our perceptual/cognitive systems.

5. REFERENCES

[1] Mathews, M.V., "The Digital Computer as a Musical Instrument," Science, Vol. 142, No. 3592, 553-557, 1963.

[2] Risset, J-C, Mathews, M.V., "Analysis of Musical-Instrument Tones," Physics Today, Vol. 22, No. 2, 23-30, 1969.

[3] Chowning, J.M., "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation," J. Audio Eng. Soc., Vol. 21, 526-534, 1973.

[4] Zwicker, E., Scharf, B., "A Model of Loudness Summation," Psychological Review, Vol. 72, 3-26, 1965.

[5] Chowning, J.M., "Perceptual Fusion and Auditory Perspective," Music, Cognition, and Computerized Sound, P.R. Cook ed., MIT Press, 261-275, 1999.