Proceedings of
the 1993 International Computer Music Conference

International Computer Music Association
and
Waseda University

September 10 – 15, 1993

Waseda University Center for Scholarly Information
and
Vario Hall, Shobi Institute of Education (Tokyo, Japan)

Sadamu Ohteru, Conference Chairman

Published by the:

International Computer Music Association
2040 Polk Street, Suite 330
San Francisco, California 94109 USA
and
ICMC 1993 Organizers
Waseda University
Tokyo, Japan

September, 1993

Computer Music: A Grand Adventure and Some Thoughts About Loudness

John M. Chowning
Center for Computer Research in Music and Acoustics
Stanford University

It was just 30 years ago that Max Mathews published an article describing his work using digital computers to synthesize and process sound. The consequences of that article for a number of fields, including music composition and theory, signal processing, auditory perception and for the music industry, have been far-reaching. We have learned much concerning the nature of sound during these years, but **loudness**, it seems, is one attribute of music that remains poorly understood with sometimes harmful ramifications.

There were a few composers in the New York area who knew of Max Mathews' work in computer sound synthesis at Bell Telephone Laboratories (BTL) in the early 1960s. But is was not until he published his article in the journal *Science* in December 1963 that I and others more distant became aware of his work[1].

The effect of this article was immediate and in some opinion seductive. It offered those of us who had access to computers (as a shared resource) a means to accomplish that which was otherwise denied. Because of the expensive equipment required, the production of electro-acoustic music, typical of the large analog studios such as the Columbia/Princeton Studio and those supported by some of the state-run broadcast entities in Europe, was beyond the means of most institutions. To many of our academic colleagues, however, we were being led astray -- following a path that would lead to trivial results in the best case and in the worst, the dehumanization of music composition. At this time I was a graduate student fortunate enough to have a perceptive and adventurous advisor, Leland Smith, who not only encouraged me to pursue the interests generated by Mathews' article but who also joined me in this work several years later.

The computer system to which I had access was powerful, indeed. It comprised an IBM 7090 that had an enormous memory of 36k 36-bit words and a hard disk whose capacity was well over 500k words and about the size of a large refrigerator. The hard disk was shared by a DEC PDP-1 computer. By September 1964, with a great amount of help from a young undergraduate mathematics major, we implemented the *Music IV* program, provided by Max, on the IBM 7090 and used the PDP-1 as buffer memory to send the samples from the hard disk to the DEC Scope. The x-y converters for deflecting the electron beam of the CRT were connected to separate channels of an audio system, thus providing for stereo output. I mention these details to remind us of the progress that has been made in three decades.

---

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

The young undergraduate who engineered this system was David Poole. He was extraordinarily helpful and patient as I, at age 30, learned from him about things that I would have never have thought to be important to music. He was not only responsible for this, the first on-line computer music system, but later went on to design and build really large computers, one of which served many years as the platform for CCRMA's digital synthesizer. This synthesizer was fondly known as the "Samson Box " after its designer Peter Samson of Systems Concepts in San Francisco.

Music IV had a very attractive surface attribute similar to the much discussed program *Max* by Miller Puckette and David Zicarelli. Processes were represented by assembling simple, well-defined units that generated and controlled acoustic signals -- signals that could reach great complexity. This representation, with its implied physics, as opposed to an explicit mathematical representation, was wonderfully accessible to those of us who were many years from our last math course.

With this powerful sound synthesis concept at the center, the adventure of computer music began. *Music IV* was so well-conceived that it was the progenitor of a number of offspring on a variety of systems[2]. These music programs were run on large batch processing or time-shared systems used by a number of people from a variety of disciplines who intermingled and conversed as they waited for jobs to run or terminals to become free. Quite by chance, surprising and serendipitous alliances were made. Seemingly disparate fields found substantive connections. Composer/musicians needed to know what engineering scientists and cognitive/perceptual psychologists knew and these same scientists needed to know what and how musicians knew and heard. Signal processing, acoustics, psychoacoustics and computer programming became familiar and necessary terrains of knowledge for musicians, and music became a rich domain of application for the engineering and perceptual sciences. Escher, Shepard, and Risset (graphic artist, cognitive psychologist, and composer/physicist) became linked through compositions in which powerful visual illusions suggested compelling auditory counterparts. Young composers were nurtured by scientists and engineers - Godfrey Winham at Princeton, David Poole and George Gucker at Stanford, Jim Beauchamp at Illinois and, of course, Max Mathews. What a beginning -- the field of music would acquire another dimension.

The beginning was not easy. In spite of the power and generality of Music IV (and its successors), the synthesis of "interesting" sounds was not easily accomplished. Computer generated sounds seemed to lack a "liveliness" that was present in sounds synthesized by analog means. One difference was due to the **precision** of the computer, a property that would turnout not to be a "curse" but rather a key to essential knowledge needed to produce sounds of great complexity and elegance.

---

[2] see S. T. Pope's excellent review of software synthesis, *Machine Tongues XV*, Computer Music Journal, Vol. 17, no. 2, (1993).

Analog synthesis, processing, and control devices in those years had electrical inaccuracies that produced small acoustic inaccuracies -- variations in frequency, or delays that smeared the signal in interesting ways. These inaccuracies seemed to engage the ear. The electrical inaccuracies of computers, however, were hidden by encoding which resulted in sounds that were dreary in their precision and rejected by our ears as being "lifeless".

The concern with the "lifelessness" of computer generated sound led to my early interest in artificial reverberation. Reverberation makes simple orderly sounds a little less so and therefore more "palatable" to the ear. Based upon the work of Manfred Schroeder at BTL, we implemented all-pass delay networks that produced dense uncolored reverberation. Reverberation also implies space, especially distance, which engendered my further interest in sound localization. Here, the attempt was to free sounds to the extent possible from their locked location at the loudspeaker sources, to create a virtual sound space. The precision of the computer allowed the coordination of several computed functions (for angle, distance and Doppler shift) that controlled the apparent motion of a sound in an illusory space. In recent years I have thought again about the relationship of this work, of nearly 30 years ago, to our understanding of loudness. I will speak about these thoughts presently, but first I will describe what I believe to be the most important investigation of these early years, an investigation that had far-reaching consequences for our understanding and ability to produce effective music using computers. It also demonstrated that psychoacoustics could be a powerful complement to good engineering.

Max Mathews and John Pierce, who was the director of the laboratory in which Max worked, are scientists having broad interests and great vision. They often made their facilities available to artists and composers. Jean-Claude Risset had been invited in 1964 to work at BTL in the area of computer music and acoustics. Risset's musical training and advanced degrees in physics combined with his extraordinary insight to produce, in 1965, a revealing investigation of trumpet tones.[3]

Using Music IV to construct additive synthesis instruments, Risset had success in synthesizing some natural instrument tones based upon data taken from books and articles on acoustics. Trumpet tones, however, eluded him: that is the acoustic knowledge of the day was insufficient to produce data that would yield an effective simulation. Risset recorded a few short trumpet tones and then performed a pitch-synchronous analysis on a computer that produced displays of the harmonic amplitudes as a function of time. From these displays Risset made several observations. The most important was that the low-order harmonics built up faster than the high-order harmonics during the attack portion of the tones, and the low-order harmonics had a longer decay. Risset then cleverly constructed an additive synthesis instrument that

[3] J.C. Risset, *A Computer Study of Trumpet Tones*, J. Acoust. Soc. Am. 38, 912, (1965). first presented at the 70th meeting of the Acoustical Society of America, Saint Louis, November, 1965. The study was also included in *Analysis of Musical-Instrument Tones,* Physic s Today, Vol. 22, No 2, Feb. (1969) with M. V. Mathews as co-author.

similarly coupled the spectral energy to the amplitude such that the centroid of the spectrum increased or decreased in frequency with an increase or decrease in overall intensity. The results of the synthesis were so successful that the difference between the original recorded tones and the synthesized tones were hardly distinguishable, even to knowledgeable subjects.

There were at least three important points to be made as a result of this investigation; 1) the use of computer analysis techniques could reveal attributes of natural tones the salience of which the ear could confirm but not itself analyze, 2) by systematically varying the parameters of a synthesized tone, judgments could be made regarding their subjective importance, thus pointing toward a psychoacoustic description rather than a purely acoustic description, and 3) one can abstract unheard detail from the psychoacoustic data describing a tone and achieve a substantial reduction in the data and an increase in computational efficiency without altering the percept.

It was with this work of Risset, I believe, that the medium of computer music at once reached a level beyond the abstract promise of the capability of producing any perceivable sound. Having acquired the knowledge and technique to synthesize natural sounding tones presupposes an understanding of the processes of nature that can find use in the creation of sounds that are conjured from one's imagination. Further, the work, accomplished in an especially supportive laboratory, affirmed the efficacy a multi-disciplinary effort, bonding the sciences and music in a relationship that has become even richer in the ensuing years.

Risset's study of trumpet tones had a major influence on my own development of FM synthesis. I first heard about this study on visit to BTL in 1967. Risset showed me his work and played some examples. I in turn showed him my initial work in FM synthesis. It was not until 1970, however, that I fully appreciated the importance of his discoveries about trumpet tones. While working on the synthesis of percussive sounds, I noted that in nearly all tones of this class the amplitude envelope and the envelope controlling the modulation index were very similar if not the identical. I wondered if there were any other class of tones where this might be the case, and I remembered Risset explaining, some three years previously, the relationship between the attack portion of the amplitude envelope and the evolution of the spectrum of a trumpet tone. Within 30 or so minutes of experiments I was able to create credible brass-like tones by simply coupling a single function to the amplitude and index envelopes with appropriate scalings. At this moment, I realized that the parameters of FM synthesis can have a remarkable perceptual relevance.[4]

This and other work by Risset and Mathews spawned much important early work in analysis-synthesis as applied to the study of timbre: Moorer and Grey, Morrill, Wessel, Rodet and Bennet, etc. This work also showed us that

[4] The ease with which spectral change can be coupled to effort (key velocity) is one of the reasons that YAMAHA's DX7 was so remarkably successful.

computers could produce sounds that were substantially more complex than those produced by analog electronic means. Many imaginative composers produced works constructed with sounds of extraordinary elegance -- sounds that possessed an auditory complexity that approached that of natural sounds. Yet even with the increasing sophistication in the understanding and control of timbre, an attribute discovered to have a number of dimensions, **loudness** has for the most part remained dependent upon the single dimension of amplitude or physical intensity. In the remainder of this address I would like to share some thoughts regarding loudness -- loudness as a multi-dimensional percept that suggests a more musical and "ear-sensitive" means of control than that of simple intensity.

Intensity is commonly thought to be the physical correlate of loudness, a subjective quantity. In order to see that loudness must include more dimensions than intensity we can imagine the following experiment.

> A listener faces two singers, one at a distance of 1 meter and the other at a distance of 50 meters. The near singer produces a **pianissimo** tone followed by the distant singer who produces a **forte** tone. Otherwise the tones have the same pitch, the same vowel, and are of the same duration. The listener is asked which of the two tones is the louder?

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 isotropic). 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 of the singer producing the **forte** tone is 50m which would result in a decrease of intensity by $1/50^2$ or 1/2500 if compared to the intensity of the same **forte** tone sung at a distance of 1m. The listener, however, is asked to judge the relative loudness where the nearer tone is a **pianissimo** rather than **forte**. Let us suppose that the intensity of the **pianissimo** is 1/1000 that of the **forte**. The greater of the two intensities reaching the listener then is the near **pianissimo**. If intensity is indeed the physical correlate of loudness, then the answer to the question "which of the two is the louder?" would be unambiguous. However, the listener's answer is that the tone at 50m is the louder even though the intensity of the nearer tone is about 2.5 times greater. How can this be so?

Spectral Cues- In the definition of the experiment it is stated that the vowel for the two tones is the same. The listener perceives the tones to be of the same timbral class: soprano tones that differ only in loudness or vocal effort. In natural sources the spectral envelope shape can change significantly as energy (or effort) changes. The spectral envelope changes shape favoring the higher component frequencies as vocal effort increases, that is, the centroid of the spectrum shifts away from the fundamental frequency.[5] Now we can begin to understand how the listener in the experiment was able to make a judgment regarding loudness that controverts intensity as the sole determinant of loudness.

Knowing the difference in timbral quality between a loudly or softly sung tone, the listener 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 intensity at the source? Again, the answer may be the distant tone even though its intensity is again the lesser of the two - <u>if</u> the experiment is performed in an environment that produces reflected energy or reverberation.

Distance, Reverberation and Auditory Perspective- The direct energy is that part of the spherical wave that passes uninterrupted, via a line of sight path, from a sound source to the listener. Reverberation is the collection of echoes, typically tens of thousands, that are reflected from the various surfaces within the space before arriving at the listener. The intensity of the reverberant energy in relation to the intensity of the direct energy aids the listener in determining that one tone is more distant and that it is louder. How does our listener in the experiment use reverberant and direct energy in this way?

If a source produces a sound having constant energy, but at increasing distances from a stationary listener, approximately the same amount of reverberant energy will arrive at the listener while the direct energy will decrease in intensity according to the inverse square law. The fact of constant energy at the source is indicated by the unchanging intensity of the reverberant energy. The listener is then able to interpret the decreasing direct energy as increasing distance rather than decreasing loudness. On the other hand, if a source produces a sound having decreasing energy, but at a constant distance, the corresponding decrease in the reverberant energy will be interpreted as a decrease in loudness. Again, the amount of energy at the source is indicated by the changing reverberant energy.

The listener in the experiment determined that the reverberant energy associated with the tone from the distant loudspeaker (singer) was greater than was the reverberant energy associated with the tone from the softly sounding near loudspeaker (singer) leading him to infer that there was greater energy at the source, therefore the tone was louder!

In vision, the change in the size of the retinal image is not normally perceived as a change in the size of an object but rather as a change in the distance of the object. This phenomenon is referred to as **size constancy** and depends upon context or perspective. In audition there is a similar phenomenon called

---

[5] Similarly, the spectral envelope shape can change significantly as pitch changes. In general, the number of partials in a spectrum decreases as pitch increases, that is, the centroid of the spectrum shifts toward the fundamental.)

loudness constancy. Again, a sound having constant energy but at increasing distances will be perceived by a listener to have constant loudness. It is the constant intensity of the reverberant energy which provides the context or *auditory* perspective.

Auditory perspective is not a metaphor in relation to visual perspective, but rather a phenomenon that seems to follow general laws of spatial perception that cross sensory modes. It is dependent upon loudness (subjective!) whose physical correlates we have seen to include spectral information and distance, in addition to intensity. Our auditory system evolved in acoustic circumstances where these dimensions of loudness are universally present. When deprived of spectral and distance information the perception of loudness becomes flat and less interesting to the auditory system.

Computers can easily be programmed to extend the dimensions of loudness beyond intensity thereby providing the composer with a control vastly more subtle, more **musical**, than that provided by intensity alone. Manufacturers of synthesizers now offer the user spectral and intensity change as a function of effort (key velocity for example). But distance as a function of constant reverberant energy in relation to a varying direct energy is still not directly possible with a modern synthesizer or sampler, yet they all include artificial reverberation.

The musical importance of these dimensions of loudness can not be over-emphasized, yet their incorporation in the music produced by means of either general purpose computers or synthesizers is not widespread. Should there be some insensitivity to the importance of these perceptual domains where listeners find their reality, perhaps these thoughts will provoke discussion. It seems unfortunate to me that with all of the knowledge that we have worked so hard to discover and understand, with all of the sophistication of our techniques and equipment, we still present concert music at levels of intensity that cause pain. If what the presenter really wants is *loudness* I hope that these considerations might suggest alternatives to simple intensity as a way of achieving it.

Finally, the issue of loudness is of general interest because it is a fundamental of music perception. The domain of sounds to which perceptual 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 perceptual issues suggests musical/acoustic events at distant boundaries that cannot be a part of our traditional musical experience, yet which can find expression through machines in ways that are consonant with our perceptual system and magical in their effect. It must be true that continued research and understanding of musical perception/cognition will ever enrich this grand adventure.