Towards Understanding the Relation between Psychological Perceptions and the Physical Synthesis of Sound

Part 1: Perceptions

Michael P. Zeleznik
Computer Music Seminar Final Project
Department of Computer Science
University of Utah

March 1982

Abstract

When science must deal with the human observer as ultimate 'quality determiner', it is often difficult to find a suitable isomorphism between the human perceptions and the particular laws by which the scientist wishes to live. Without this, measurement and hence understanding is severely limited. Such a problem currently exists in the field of music synthesis, where auditory perceptions and mathematical methods of sound generation have not yet found a common ground. This paper in multiple parts, will explore this problem, first discussing various psychological aspects of sound and music perception, then presenting some methods used to synthesize sounds from existing musical instruments. One attempt at bridging the apparent gap will be discussed. Part 1 deals with the psychological aspects. Part 2 will deal with current methods of synthesis. Part 3 will attempt to give insight into bridging the gap.

* predictable results . . .

Table of Contents

In 1982 I was quite verbose and rambling.
If you can get past that, the core content and ideas are still valid.

1 The Problem

In order to produce something, a definition of how to do it must exist. This may sound trivial, but from just where does such a definition come? Surely it is best to come after the specifications of the final product have been determined. From these specifications, the required equipment, raw materials and method of production can be determined. This is not earth shaking, but it is very important.

If one wished to synthesize a field of grass, how would it be done? It would certainly depend on the intended use of the field. If it were to be a football field, the grass had better be durable and the base provide good footing for spikes (and probably be green), though it need not be edible and probably should not grow. If, however, the field was for grazing cattle, it had better better be edible and should provide for replacing itself (once ate), but it need not be too durable and footing is not an issue. The point is that in order to synthesize something, the important qualities of the end product (output) must be determined; not all of the qualities; just the ones important for the particular application.

Knowing the important grass qualities, a method of producing it is then determined. The method of synthesis has followed from the important qualities (the definition) of the output. Thus, the parameters of the synthesizing process are guaranteed to be

closely related to the qualities of the output. This isomorphism is a result of how the problem was approached, and makes the entire situation quite nice. On the other hand, if a grass manufacturing plant had been set up, a priori, to produce grass made of plastic, it would be a major effort to synthesize the grazing field. The isomorphism between the definition of the output and the method of synthesis has disappeared. The required output qualities are incompatible with what the synthesizer can produce, and this makes the entire situation quite disconcerting.

In sound synthesis the problem of obtaining a reasonable relationship between the output (the sound) qualities and the synthesizer input parameters becomes instantly compounded since the human observer is now used to define the output qualities. The same problem exists in any field that uses the human as observer; for example, image processing. In general, the methods of synthesising sound are based on some form of mathematics, and there is no reason to believe that the laws of mathematics will be very useful in understanding the psychological behavior of the human being. These laws were contrived precisely to describe our daily encounters with the external, physical world; it should therefore come as no surprise that they work there. Actually, mathematical analysis, in all but the simplest cases serves only to approximate its behavior, and often poorly. So why even entertain the notion that this will work in understanding our minds. For a book on the skills of mathematics applied to

psychology, see "Quantitative Methods in Psychology" by D. Lewis, 1948, Iowa City [9]. If the sound output were being monitored by a device which simply measured sound wave amplitude, and that were the only important quality, then the synthesizer output could be turned up or down until the meter indicated the desired amplitude. Since both the analyzer and the synthesizer obey the same laws of mathematics, the isomorphism exists. When, however, the human listener indicates that the sound is too warm, or that the 'feeling' is too open, which synthesizer knob should be changed to correct those qualities? Probably none, since the thing was never designed for those input parameters; back to the plastic grass plant again! How do we relate the subjective, psychological sensations of sounds and music, to some type of objective, physical properties of sound waves, so that they can be produced.

2 What do we hear?

Without going into the anatomy and physiology of the auditory system (for an excellent treatment of this subject, see Yost (1977) or Coren (1978)), how does one discuss what is heard when listening to what is generally agreed upon to be musical tones? Yes, already a definition has been assumed before the fact. Actually, if the qualities about to be discussed are present when listening to some auditory stimulus, then it shall be considered to be a musical tone; if not then it isn't. This is an over simplification of a nontrivial problem of definition, but good enough for now. Back to the point; what do we hear, or more

precisely, how do we interpret and express what we hear?

In order to answer this question in any meaningful way, an organized approach to the study of such psychological behavior is required. Such endeavors have come under the heading of psychophysics. Just like physics, psychophysics attempts to make predictions on the evolution of a system subjected to certain initial conditions. The system is the brain and associated peripherals, the initial conditions are the sensory input stimuli, and the evolution is the physiological reactions or the 'behavior' of that body and brain system, after the stimuli [8]. It should be noted that the laws of physics were devised to describe the physical world in which we live. The laws of psychophysics must describe the behavior of an individual as he interacts with that physical world. I see no reason to believe, a priori, that any similarity should exist between these two sets of laws.

For a long time, it was believed that a one-to-one correspondence existed between such psychophysical variables and the physically measurable aspects of the stimuli; it would certainly have made things simpler! Loudness would relate to sound wave amplitude and pitch to frequency. But why should that be? What if the measuring devices had instead measured some other parameters of the occurrence? Would loudness and pitch be directly related to these? The methods of measurement are a

product of the physical occurrence and the mathematics created to describe it. Human perceptions need know nothing of this mathematics, so why expect a relationship? For example, it is easy to measure the amplitude of a signal as a function of time, but it is often much easier to think in frequency space (much more difficult to measure directly). Mathematical transforms relate the two spaces. Or, though we like to think in miles-per-gallon, it is much easier to measure this indirectly and then convert. What we think in and what we measure are rarely the same.

In general agreement among all cultures, the notions of pitch, loudness and quality (timbre) seem to be fundamental measures of our interpretations of tones. With minimal, non-precise definitions of these, listeners can readily express their perceptions of tones in these terms. With a little thought, a few more notions can be added, though it may be argued that these can all be expressed as combinations of the above basis triple. These are volume (the sense in which space is filled or the sound seems large), density (the compactness or hardness of the sound), location and duration. Even more complex and subjective are the attributes of consonance and dissonance (how two tones 'go together' or 'clash'). First, we look at the results of some experiments performed with extremely simple stimuli.

2.1 Detection of sound: was there one?

Some of the simplest types of experiments dealing with human perceptions are those looking for the minimum detectable experience of a stimuli; in this case the minimum auditory experience. Is there or isn't there a sound? Or, more precisely, is there a difference from what was there before? These are often termed JND (Just noticeable difference) experiments.

2.1.1 Some results from this type of experiment

The physical variable that most closely predicts out perception of loudness is sound pressure level. The minimum audible field , that is, the minimum sound pressure as measured in free field (as opposed to at the ear drum - minimum audible pressure) required for one to just discern that a sound is present, is a function of frequency, the ear being most sensitive to those frequencies between 2000 Hz and 5000 Hz. We are about 100 times less sensitive to a 100 Hz signal than to a 3000 Hz one, which is roughly the peak of sensitivity. Within this range, the ear has a dynamic range of up to 150 dB (7.5 million- to-one change in sound pressure), though this is certainly well into the threshold of pain region!

Sound pressure is not the only factor involved with the minimum audible field. The duration of the stimulus comes into play. It appears that some sort of minimum sound energy is required in order to perceive a sound. For durations less than 250-500 msec, the approximate relationship between the power (P) and duration

(T) of a just perceivable sound is P x T, which is the definition of the energy of the sound. This is somewhat frequency dependent. Beyond 250-500 msec, additional duration does not increase detectability [11].

Also, by increasing the number of frequencies presented together, the likelihood of hearing the sound is increased. For frequencies near to each other, it is as if the energies sum (eg. two simultaneous tones, each being half of the required intensity to be heard alone, will be detected). For frequencies far removed from each other, this does not hold. There is a type of critical bandwidth of frequencies, beyond which adding tones does not increase detectability, just as with the 250-500 msec duration limit above. This bandwidth varied with frequency, increasing at the higher ones [5].

Additionally, minimum audible field thresholds for two ear stimulation are about half that for one ear stimulation, and the two stimuli need not even occur simultaneously. If the tones are presented separately to each ear, with the total combined duration less than about 200 msec, detection will occur with each tone roughly half the level required for one ear detection. Two ear stimulation also increases the subjective impression of loudness in addition to simply lowering the threshold of detectability. This may indicate that the summation of information happens in a somewhat central region of the brain, as

the separate information does not meet until late in the neural pathways [5].

Another area of interest lies in the ability of some tones to mask others, the whole process being rather selective. Generally, increasing the masking tone level necessitates an increase in the target tone level to maintain detection of the target. The most strongly masked tones have frequencies very close to the masking tone. Lower tones mask higher ones better than the reverse, though the effect diminishes as the frequency separation increases. Some explanations for this have been based on the physics of the ear, but this can not explain the masking effects of different tones to the separate ears. In this case, the masker intensity needs to be increased and the effects seem more symmetric with respect to frequency. One seemingly anomalous result was obtained when white noise was presented to both ears (in phase) with a target tone to only one. The target was much more detectable than with noise and target only presented to one ear [5].

In discussing our ability to discriminate between two stimuli, a ratio called the Weber fraction (W) is often utilized. This is simply the ratio of the amount of change just required to be perceived (dS), to the standard stimulus from which the change occurred (S), W = dS / S. For sound intensity discrimination, the Weber fraction is a function of both frequency and intensity,

though it is relatively constant over a wide range of intensities. We are most sensitive to changes in the mid-frequency range (2000 - 5000 Hz), being sensitive enough to reliably detect a 20% change over the broad range of frequencies and intensities where most of our everyday hearing takes place [11] [5].

As for frequency discrimination based on JND experiments, it is seen that the Weber fraction is nearly constant (.01 - .005) for frequencies above 1000 Hz, and is more consistent at higher sensation levels (eg. a 60 dB signal maintains a Weber fraction below .01 for all frequencies above 250 Hz, while a 10 dB signal maintains this only above 500 Hz). For example, at a frequency of 400 Hz, a change of 2 Hz can be detected. Below 200 - 500 Hz, this fraction increases quite rapidly, rising to .04 - .07 in the 60 - 100 Hz range [11].

2.1.2 What does this all mean?

It should be obvious that even with simple experiments presenting single tones in absolutely controlled environments, the data are not always easy to interpret for an individual experiment. Attempting to tie it all together is even more frustrating. Certainly, in the context in which the experiment was carried out, the data is quite meaningful, but trying to extrapolate into other environments is a risky proposition.

For those familiar with classical and quantum physics, this

should be obvious. At the turn of the century, classical physics was a relatively concise set of mathematical theories which did a reasonable job at predicting most of the then encountered physical phenomena. The attempt to extend those theories to the atomic level resulted in an entirely new science, quantum physics. Classical physics ,in the end, is nothing more than a special case of quantum physics. Just as quantum physics insists that the measurement process perturbs the system, so do our experiments perturb both the psychological state of the observer and the physical state of the sound. How often have you listened to coherent white noise through both ears with a pure sine wave into one? And, Just as attempting to extend classical physics into the realm of the atom served to demonstrate its inadequacies, so may attempting to extend the results of such simple experiments to the realm of real music. I do not intend to underplay these results, or to minimize the importance of such experiments; only to caution.

An example of a similar problem in the image processing field is apparent in the recently completed work of H. Ravindra at the University of Utah (Computer Science Department), dealing with the human visual system. He has succeeded in building a mathematical model of the visual system that correctly predicts JND results for human observers of solid shaded circular objects in the center of a rectangular uniform field (black and white only). How this data relates to any more complex image structure

is unknown, and is probably another doctoral thesis in itself.

2.2 Dimensions of sound: What does it sound like?

Discerning whether there was or wasn't a sound is all well and good, but the real question is what notions did that sound evoke. This proves to be a much more difficult area to explore, mainly (my opinion) because we are not sure how to formulate the questions. Don't misunderstand; I have no answers! Now observe some results of experiments attempting to quantify or even just categorize the responses which deal with loudness, pitch, timbre, volume, density and general subjective impressions of sound and music.

2.2.1 Loudness

Our impressions of loudness are not directly related to stimulus intensity. Decibels are not a measure of loudness. It has been shown that loudness (L) varies as a function of signal intensity (I) in dB, as L = aI**e , where a is a constant. The exponent e actually depends on the specific stimuli and the test conditions, but one text cites a value of e = 2/3 [5]. Because of the frequency dependence of loudness, both due to our own frequency dependent thresholds as well as to other psychological factors, special units have been adopted (phons and sones) to divide loudness levels into equal increments according to the hearing mechanism [1].

Besides the frequency dependence of loudness, tonal duration

also has an effect similar to that in the minimal audible field experiments. Shorter tones (below about 200 msec) require larger intensities to be perceived as loud as a longer tone. However, a reverse effect also exists, in that the longer a continuous tone is presented, the less loud it will seem. This is termed auditory adaptation, and should not be confused with a similar phenomenon called auditory fatigue, where the general presense of all sounds tends to reduce the overall sensitivity of the hearing process for extended periods after the exposure. This is why the clock radio or car radio often seems so loud in the morning. Another influential factor is the complexity of the stimulus. Listeners asked to match the loudness of a pure tone with a complex tone have shown that, for complexity exceeding some lower bandwidth limit, apparent loudness increases with increasing complexity bandwidth though the overall energy is the same. Since this is the normal setting for music listening (ie. non-pure tones), it is difficult to say what the pure tone tests indicate in a practical sense.

In the search for a single underlying physiological variable that may be responsible for the concept of loudness (one that is somehow affected by all of the physical ones above), it has been suggested that loudness is solely determined by the total amount of neural activity per unit time in some area of the brain. Though this may be the case, it is not entirely supported by the evidence [5].

2.2.2 Pitch

One of the more deceiving notions of quality is pitch. Originally thought to be related directly to the waveform frequency, and then to the lowest frequency present (fundamental), it is now known that, depending on the make-up of the wave, perceived pitch can be equivalent to that of a pure tone whose frequency is far below any frequency present in the waveform. It is as though the frequency has been sub-synthesized. Also, just as frequency affected apparent loudness, intensity affects pitch. The higher the intensity, the higher the perceived pitch. Tone duration, as it has in nearly every experiment, affects pitch also, the critical duration again in the 250 msec range. Below this time, increasing the tonal duration improves the ability to differentiate different pitched tones. The minimum length of any tone to be perceived as having a pitch (as opposed to being a click) is about 10 msec, though lower frequency tones require that on the order of 10 cycles reach the ear before pitch is perceived. For a 50 Hz tone, this is 20 msec.

One fundamental difference between hearing and seeing lies in the ability of the hearing mechanism to apparently separate a waveform into its Fourier components in certain situations, allowing us to perceive them separately, as was investigated by Helmholtz in the mid 1800's. If two different tones are sounded simultaneously, the listener can distinguish the two, even though

the waveform is the superposition of them. In the eye this is not the case (eg. red and green light rays together are 'seen' as yellow). Some people can distinguish up to 6 or 7 separate harmonics. This is known as Ohm's Acoustical Law.

In short, pitch perception is one of the more elusive aspects of our hearing mechanism. Much research is being done, but much controversy still exists.

2.2.3 Timbre

Timbre, in a broad sense (it is not well defined) can mean all of the tonal qualities that characterize a particular musical sound; a broad definition indeed! Though many attempts at defining this elusive term have been mounted, there is no formal agreement on the meaning of this term in relation to the auditory phenomenon which should be included in its definition [6]. It appears to be, however, that quality most closely related to the harmonic content of a tone, though many other factors come into play. In 1863, Helmholtz (in one publication), summarized various 'feelings' as related to the harmonic content of a tone. For example, the fundamental alone was termed 'soft'; the fundamental plus first harmonic was 'mellow'; the fundamental plus higher harmonics was 'sharp'; overpowering harmonics with a less intense fundamental was 'hollow'; dominating odd harmonics was 'nasal'; and so on [5].

Timbre perception is not a well understood subject in the field

of auditory perception, due to its vast complexity [6]. Little research has been directed towards it. Such sound waves as pure tones, pulse trains, and small sets of sine waves have received more attention as they are easier to characterize and produce, and simple enough to allow reasonable interpretations of the listeners responses. Even of the little research devoted to timbre, much of it has restricted the analysis to that of steady- state periodic waveforms, thereby neglecting the temporal variations occurring in all natural phenomena (normal listening environments). This was clearly the case with Helmholtz.

Many psychological studies on verbal representation of attributes of timbre have found three to be common; brightness, fullness and roughness. Brightness is associated with the frequency of the midpoint of the energy distribution. Fullness is some function of the relative presence of odd or even harmonics. Roughness is associated with the presense of consecutive higher harmonics above the 6th and is a function of the location of these harmonics (a correspondence to dissonance). It is also accepted that the attack and decay of the tone, and more importantly of the individual harmonics (since each may be different temporally), are essential for recognition of instrument sounds.

I can only serve to indicate the vast scope of this concept in this paper. It should be at least obvious that it is not a

trivial matter. For a much more detailed and interesting study of musical timbre, refer to Grey (1975).

2.3 Really subjective impressions: The tuff ones

Until this point, we have only dealt with gross abstractions of the real world of sound and music. Suppose one wished to write the sound track for a movie scene consisting of a panoramic view of outer space with far off stars and galaxies. One would probably choose a full orchestra with large booming drums over a banjo and piccolo. Subjectively, some sounds evoke a feeling of 'vastness'. Some sounds are more 'expansive' or space filling than others. This attribute is termed volume. Density is a much less tangible concept, indicating the relative tightness, concentration or hardness of a tone. It seems to operate as the inverse of volume, with higher pitched or higher intensity sounds appearing denser.

As can be seen, the precise definitions of sound attributes have given way to imprecise generalities. To carry this further, imagine the voices for a cartoon skit with an ogre and a captured princess. Without much doubt, the ogre's voice would be low in pitch and slow spoken. The princess would probably speak fast and high. It surely may be argued that these are conditioned responses, but that is not the issue. The issue is how to quantify and understand the response. In fact, these characteristic meanings of low/slow and high/fast tones are readily observable in many animals, where the type of sound

(characterized by its predominant frequency and tonal durations) emitted can be related to the animals meaning. Based on data from monkeys, for example, this can be traced from a squeak, characteristic of a defeated, exhausted animal, to a roar, characteristic of a confident, threatening animal [5]. In another example, observe that the higher frequency notes occur higher up on the stave.

Though the previous examples may not seem relevant to music synthesis, many questions of the same type can be very important. Consider the work done by Bose in the early 70's [3] [4]. In an attempt to determine the important qualities of reproduced sound (partly looking for the cause of 'shrillness' in music), much progress was made. As it was, 'shrillness' was not predictable from any of the normally measured parameters of sound. It depended on the characteristics of the listening environment (the field reverberance). The proof of the result is evident to anyone who has listened to a pair of Bose speakers. The papers referenced provide a very nice discussion of sound reproduction issues.

As a parting comment, consider how to define and measure the subjective magnitude that represents the subjective magnitude that represents the urge to bring a given melody to its tonic completion.

3 Conclusion

In a presentation of this length, it is difficult to present a thorough overview of the work done so far in the area of understanding auditory perception. I have tried to discuss the generally accepted, key items, somewhat common to much of the literature. We are certainly a long way from understanding the process in a way that is compatible with most current methods of synthesis. An overview of these methods will be presented in the second part of this article. It is hoped that at some time, a more common framework can be constructed, in which the language of perceptions can be more easily related to the language of the physical synthesis techniques employed at that time. Ideally, the synthesis techniques would derive from the perceptual analysis.

REFERENCES

[1] Wayne Bateman.
Introduction to Computer Music.
Wiley-Interscience, 1980.

[2] Paul C. Boomsliter, W. Creel.
Research Potentials in Auditory Characteristics of Violin Tone.
1969.
77th Convention - Acoustical Society of America: Symposium on Violin Acoustics, April 11, 1969.

[3] Amar G. Bose, T. Stockham Jr.
Sound Recording and Reproduction - Part One: Devices, Measurements , and Perception.
Technology Review :19-25, June, 1973.

[4] Amar G. Bose, T. Stockham Jr.
Sound Recording and Reproduction - Part Two: Spatial and Temporal Dimensions.
Technology Review :25-33, July/August, 1793.

[5] Stanley Coren.
Sensation and Perception.
Academic Press, 1979.

[6] John M. Grey.
An Exploration of Musical Timbre.
Technical Report STAN-M-2, Stanford University, Department of Music, February, 1975.

[7] Richard C. Heyser.
The Delay Plane, Objective Analysis of Subjective Properties: Part I.
Journal of the Audio Engineering Society 21(9):690-701, November, 1973.

[8] Juan G. Roederer.
Introduction to the Physics and Psychophysics of Music.
Springer-Verlag, 1973.

[9] S. S. Stevens.
Handbook of Experimental Psychology.
John Wiley & Sons, 1965.

[10] Fritz Winckel.
Music, Sound and Sensation - a modern exposition.
Dover Publications, 1967.

[11] William A. Yost.
Fundamentals of Hearing.
Holt, Rinehart and Winston, 1977.