PSYCHOACOUSTICS AND CONTEMPORARY MUSIC

W. M. Hartmann

Physics Department, Michigan State University, East Lansing, MI 48824 USA and IRCAM, 31 rue Saint Merri, 75004 Paris, France

It is evident that the nature of the human auditory system has profoundly influenced the parameters of our music. One has only to imagine how different, in physical terms, music would be if the human auditory system resembled that of the bat or that of the whale. In these cases meaningful music would most likely consist of a rich array of glissandi, respectively scaled to ultrasonic frequencies, with emphasis placed upon a rather narrow band, or scaled to very low frequencies.

But our ears are as they are. Their spectral and temporal characteristics are matched by our human voices and by our human musical instruments and compositions. Our bass instruments are powerful; they need to radiate sounds of relatively large intensity, in part, because of the relative insensitivity of the auditory system at low frequencies. The limits of frequency analysis by the auditory system have placed corresponding limits on musical composition. Composers traditionally avoid small intervals, thirds and smaller, in the deep bass. Low tones separated by small intervals mutually interfere with a jarring roughness because the auditory system is unable to resolve them.

There is thus a common bond between the psychoacoustician, whose goal is a first-principles understanding of the auditory system, and the creator of contemporary music. Both are explorers in a territory which is only roughly charted; they are natural travelling companions, though they have different backgrounds and different views of the purpose of the journey. The psychoacoustician is typically interested in various limits of the auditory system, probed by diverse extreme acoustical combinations sequences and contexts. Contemporary music explores such limits, sometimes in almost scientific simplicity, often as imaginative variations on psychoacoustical themes or competitions between perceptual effects. The contemporary musician understands that ultimate acceptance of his music depends upon appeal to the auditory system and its complex of bi-directional interactions with cognitive processes. He is naturally curious about what has been learned scientifically about this system.

One theme recurs repeatedly throughout psychoacoustics: the ear is a frequency analyzer. From the inner ear to the auditory cortex the auditory system is designed in such a way as to process sounds in channels distinguished by frequency. These are channels of information transmission from the acoustical world to the perception of that world. Cues relating to the detection of a sound in a background, cues for pitch, for localization, for timbre and source recognition, on articulation and expression are all accumulated by adding information from the individual channels. The spectral balance of the traditional orchestra is one, among many, which makes good use of all of the frequency channels.

Music confined to a small frequency region deprives the listener of information. Spectral imbalance may be used musically for contrast. For example, a shimmering high-pitched chorus, in isolation or separated by a wide spectral gap from deep bass, conveys the image of wonderous action, the more mysterious as its details are dimly perceived. Indeed, because of the rich harmonic content of musical instruments, for most combinations of instruments, spectral imbalance is probably no more severe than other restrictions which might be imposed, such as restrictions in dynamics or tempo or key. However, in electronic music, where extreme cases can be realized, or with some instrument combinations spectral imbalance may produce unexpected consequences. Recorder ensembles induce difference tones in the non-linear human ear, which are not masked by any bass. As luck would have it musical thirds are notoriously efficient sources of difference tones. The difference tones may be musically acceptable if they were anticipated in the composition, but extended listening to an ensemble of these supposedly gentile instruments can make one's ears ring as as surely as a rock concert. The piccolo solo is another extreme example. When listening to brief, intense high notes with little harmonic content one is often aware of a thump accompanying each note, as though the acoustic signal were rectified by the ear and the transient DC offset added to the original. The origin of the effect is uncertain, but the slight delay between the tone onset and the thump immediately suggests the acoustic reflex of the middle ear muscles. Such effects do not occur in a spectrally balanced context.

One psychoacoustical technique has taught us more about the auditory system during the past 40 years than any other: that is the masking experiment, one in which an intense sound causes a weaker sound to disappear altogether. There are masking effects in music, but the most important of these do not actually result in the total disappearance of musical tones. Instead, mutual masking effects among instruments in an ensemble alter the characters of the individual sounds. The high harmonics of bass tones, such as the double bass, cello and bassoon, are masked or partially masked by the more intense fundamental components of the mid-range instruments. As a result bass instruments have a more mellow timbre in an ensemble and sound more brilliant in solo.

Masking effects obscure transient information which comes from the onset of an instrumental tone or from small fluctuations during sustained tones. Because masking effects depend strongly upon the frequency separation between components, they are important in close harmony and much less important in open harmony. Mutual masking of transients is probably the most important difference between solo playing and unison playing.

The transients in instrumental tones play a role which is remarkably like that of the transients in speech. The parallel is worth pursuing. In speech most of the energy resides in the fixed vowel sounds; most of the information content is found in the transient consonant sounds and vowel formant transitions. The transients in a musical tone are similarly weak in intensity compared to the steady tone, but the amount of information conveyed by transients in a musical sound is enormous. Attack information or transient fluctuations within a tone are essential for musical instrument recognition. The converse is that speech is understandable and musical instruments are recognizable under conditions of extreme filtering which cause the spectral profile of sustained sounds to bear almost no resemblance to the original. Transients in musical tones contain the expressive quality of the individual note. They essentially inform the listener that a human performer is creating a tone in a particular manner on a particular instrument. By its high information content the solo calls attention to itself. One does not use a solo instrument for background music, for the same reason that one does not use speech as background. The idiosyncratic transients of the solo voice impart to it the personality of an individual. The mutual masking of these individual transients in unison playing recalls the power of a group unified in purpose.

A practical proof of the mutual masking of transients in ensemble is an everyday experience in the synthesis of electronic music. It is enormously more difficult to synthesize a realistic solo passage than it is to synthesize a convincing chorus. To synthesize the solo, one must approximate the complexity and the variety of authentic transients. Failing that the solo is rejected by the listener as boring, without soul, without even spine. To synthesize a unison chorus one can omit all the details of the transients in favor of a generalized aleatory action. The listener is not unsatisfied. He never expected to hear the details.

Finally one must note that masking effects are not linear. As a result the mutual interactions among instruments are different for different dynamic levels. Generally the relative amount of masking increases as the overall level increases, so that, for example, forte playing should enhance choral effects.

A most remarkable characteristic of the auditory system is its capacity to combine and to separate sounds into auditory images. Consider the acoustics of the string quartet. Pressure waves produced by the individual instruments are simply added together, apparently hopelessly entangled in one waveform. To this waveform is added the early reflections from the walls of the room, and finally the reverberation is superimposed. From this composit the human ear is able to extract the individual voices and to recreate the images of the instruments of the quartet. Moreover, the ear is able to make exquisite judgements concerning the intonation and articulation of each of the individual voices . . . a most remarkable feat of information processing indeed! Contemporary musicians have experimented with auditory image formation, causing multiple tones to fuse and single tones to separate. Interest in the psychoacoustics of image formation has recently been growing.

There are a number of aspects of image formation to consider. First is the combination of the harmonics of a single source into a single acoustical image. Because the auditory system initially segregates the harmonics by frequency channel the recombination must represent a subsequent stage of processing, most probably a data reduction process, which relieves the brain of the task of coping with a plethora of resolved harmonics. By modulating one harmonic or a group of harmonics in some way one can cause the single image to split. Then there is the process of distinguishing an original sound from its reflected versions. The role that this process plays in localization of the sound, the precedence effect, indicates that the process must be very rapid. Of necessity it must rely heavily on peripheral processing where the initial frequency analysis is preeminant. Next, the process of identifying individual voices is a complex process of pattern fitting. Discrepancies in onset times small enough to be acceptable to any conductor are still large enough to provide a potent cue. Random and intentional modulation in the individual sources is another useable aspect of the pattern.

A related phenomenon is stream segregation or stream fusion, whereby a series of notes appears to separate perceptually into several groups or to maintain a coherence as a time-ordered series. Most studies of the effect have used frequency differences among the notes of the series and presentation rate as parameters. Large frequency differences and rapid rate promote segregation. But some experiments have shown that streams can be segregated by timbre as well, raising the spectre that the process may be a rather cognitive one. Experiments done at IRCAM, however, have shown that the segregation by timbre is based upon the brightness dimension, so that the auditory periphery may, in fact, perform the bulk of the task, a view consistent with our view of the precedence effect.

Examples of music in which stream segregation is deliberately induced abound in the old and in the new musical literature. The psychoacoustician finds much to learn from each, about this and other topics. But it is in new music where the psychoacoustician and the musician have the opportunity to collaborate on the most fundamental level with the best chances for an effective synergism.

Acknowledgements: I wish to acknowledge helpful discussions with Pierre Boulez, Stephen McAdams, and David Wessel.