Invited talk at the
First International Conference on Music Perception and Cognition
Kyoto, Japan
17-19 October 1989
Proceedings of the First International Conference on Music Perception and Cognition pp 299-304.

AUDITORY GROUPING AND THE AUDITORY PERIPHERY

William Morris Hartmann
Physics Department, Michigan State University
East Lansing, Michigan, 48824, USA

Abstract

When one listens to music, several grouping processes take place in the auditory system. One of them is simultaneous grouping, where the spectral components from a single acoustical source are collected together by the auditory system to form a source image or auditory entity. In the case of polyphonic music, the components from a second source, possibly interleaved with the components from the first, are collected together to form a second entity, distinct from the first. In addition to simultaneous grouping, there is sequential grouping, where a series of tones is parsed into separate melodic lines, often associated with separate auditory entities.

The human ability to accomplish these groupings can be studied under simplified and controlled conditions in the psychoacoustical laboratory. Well-controlled stimuli in combination with models of processing at the auditory periphery enable one to make informed guesses about the kind of neural information that is available to higher auditory centers. This is the information on which grouping must be based. Sometimes models of peripheral processing are highly suggestive about what might be important to the higher centers.

Simultaneous grouping has been studied with the mistuned harmonic experiment, where one partial of a complex tone is mistuned from its correct value. Of interest is ability of a listener to hear the mistuned partial as a separate auditory entity, as a function of the amount of mistuning and as a function of the tone duration. The data have curious non-monotonic dependences, and these suggest that higher centers perform a kind of autocorrelation analysis, within tuned peripheral channels. An alternative model is one that is based upon the interspike interval histogram as observed in eighth-nerve fibers. This model has the advantage of including certain phase dependences that are observed experimentally; it also allows one to understand both the detection of the mistuned harmonic and the pitch of the mistuned harmonic from a single model.

The matter of sequential grouping has been studied with experiments on stream segregation. Of interest here is the conjecture that the ability to segregate auditory streams can be understood very simply in terms of peripheral channeling. Only two kinds of peripheral channel are allowed in this model, lateral and tonotopic. This idea

has been tested against experiments on interleaved melody identification, where the two interleaved melodies differed in a dozen different ways. Although the peripheral-channeling model is evidently incomplete, it can account for a large amount of data and may well indicate the most important single aspect of sequential grouping.

INTRODUCTION

When one listens to music, the auditory system performs several grouping processes, as described by Moore (1982) and by McAdams (1984). One of these is simultaneous grouping, where the spectral components from a single acoustical source are integrated by the system to form a source image or auditory entity. The other is sequential grouping, where a series of tones is parsed into separate melodic lines, often associated with separate auditory entities.

The human ability to accomplish these groupings can be studied under simplified and controlled conditions in the psychoacoustical laboratory. Well-controlled stimuli in combination with models of processing at the auditory periphery enable one to make informed guesses about the kind of neural information that is available to higher auditory centers. This is the information on which grouping must be based. Sometimes models of peripheral processing are highly suggestive about what might be important to the higher centers.

SIMULTANEOUS GROUPING

In the case of polyphonic music it is common for a listener to be able to hear out the individual voices. Viewed in the light of traditional auditory theory this ability is remarkable indeed because the theory emphasizes tuning in the auditory system, namely the separation of different frequency regions into physiologically distinct channels. But the spectral components of different voices in polyphonic music generally overlap. Therefore, it frequently happens that each tuned channel in the auditory system has some components from each of several voices. Nonetheless the auditory system copes somehow. Exactly how is an interesting question. A survey of the possibilities was given in a recent review (Hartmann, 1988).

Of primary importance is probably the fact that the onsets of different voices in polyphonic music are not precisely simultaneous. This effect has been studied by Rasch (1978,79) who found asynchrony among onsets of tens of milliseconds in performed music where the score called for simultaneous onsets. Asynchrony of this order leads to increased transparency in the music, an ability to hear individual instruments, while still maintaining the impression that the musicians are playing correctly with respect to the beat.

In the absence of onset asynchrony or differences in tonal microstructure, slight deviations in tuning can provide important cues to grouping. It appears that a listener is able to group together the components that are exact harmonics of a fundamental and segregate these from the components that are exact harmonics of a different fundamental. Perhaps the simplest non-trivial example of this is the mistuned harmonic experiment where one harmonic of a complex periodic tone is mistuned from its correct value. Of interest is ability of a listener to hear the mistuned partial as a separate auditory entity. For instance, if one begins with a complex tone with a fundamental frequency of 200 Hz and then mistunes the

fourth harmonic, the listener may experience the impression of a flute-like tone (near 800 Hz) emerging from a bassoon-like tone (200 Hz). Varying the parameters of the experimental tone can lead to considerable insight into the auditory processes involved.

Duration effects: If the duration is very short (e.g. 10 ms) then it is impossible to recognize a mistuned harmonic. This illustrates the general rule that brief duration favors integration. Any sound, no matter how inharmonic, will be heard as a single entity if is is made short enough. A second point is that segregation, as in the identification of a mistuned harmonic, requires information. Integration is the default operation.

As one increases the duration of the tone it becomes increasingly possible to hear out a mistuned harmonic. This can be shown by a discrimination experiment, as performed by Moore et al. (1985, 1986) or by Hartmann (1985). Here the listener hears two tones. One of them is perfectly periodic; the other is the same except that one harmonic has been mistuned. The listener's task in this two-interval forced-choice task is to say which tone has the inharmonic partial. It is usual to randomize the fundamental frequency so that the listener cannot use his sense of pitch to do the task. Data from such an experiment are shown in Fig. 1. Here there were seven equal-amplitude harmonics of a 200 Hz fundamental with the 4th harmonic mistuned by 2.5% (20 Hz). The sound pressure level was 40 dB. The data show that as the tone duration increased the performance, as measured by d', increased as expected. But closer inspection shows a peculiar effect, a plateau where the performance does not improve as the duration increases from 40 ms to 60 ms. Because the error bars are rather large one might be inclined to ignore this effect, but it occurs in the data for other listeners as well; for some the flat region near 50 ms actually becomes a valley.

These data suggest that the mistuned harmonic is detected by some form of neural timing process that resembles autocorrelation. The mistuning disrupts neural synchrony and leads to a reduced autocorrelation. As the duration increases from zero it becomes increasingly possible for a central processor to find time intervals over which the autocorrelation is small. However, as the duration approaches 50 ms (the reciprocal of 20 Hz) the trend does not continue. Autocorrelations over the longest available time intervals are large again because the signal resembles its initial form.

Frequency effects: As the mistuning of a harmonic increases a listener should find that it is increasingly easy to hear out that harmonic as a separate entity. However, if there is any merit to the notion that a form of neural autocorrelator is responsible for plateaus or valleys in a time function like Fig. 1 then there ought as well to be plateaus or valleys as a function of mistuning. Fig. 2 shows the performance as the mistuning of the 4th harmonic of a 200 Hz fundamental is increased. Because the tone duration was 30 ms the autocorrelator principle predicts that there should be a plateau in the vicinity of 33 Hz ms. The data in Fig. 2 seem to agree. Similarly we found that for a tone duration of 50 ms there was a plateau in performance when mistuning approached 20 Hz.

Tuning effects: To do the mistuned harmonic discrimination task requires that a listener somehow compare neural representations of the different harmonics of the tone. For instance, the duration and frequency shift dependences above suggest a comparison based

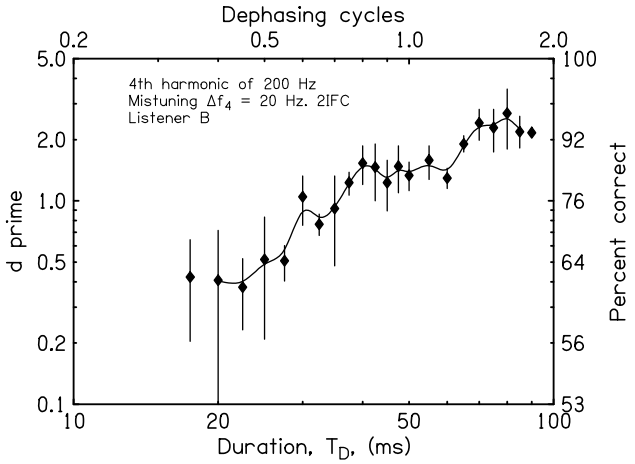

Figure 1: The ability of listener B to discriminate between a tone with a mistuned 4th harmonic and a perfectly periodic tone as a function of the tone duration. The mistuning (frequency shift Δf_4) was 20 Hz and the level was 40 dB SPL. Error bars are two standard deviations in overall length.

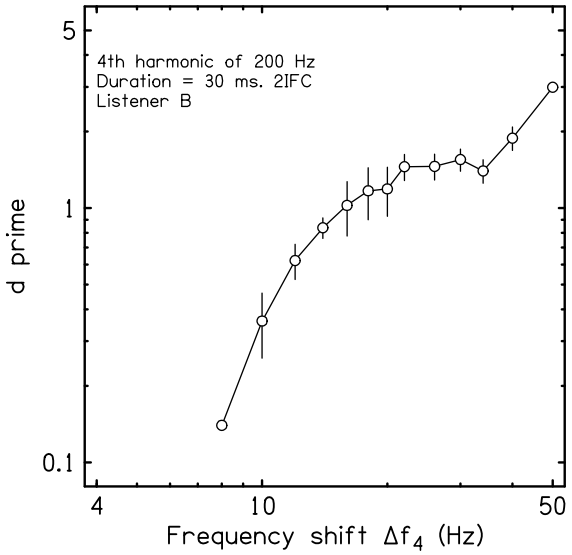

Figure 2: The ability of listener B to discriminate between a tone with a mistuned 4th harmonic and a perfectly periodic tone as a function of the mistuning. The tone duration (T_D) was 30 ms and the level was 40 dB SPL. Error bars are two standard deviations in overall length.

upon autocorrelation. One wonders how this comparison is done. Does the system compare the timing pattern of each harmonic with the timing pattern of a neighboring harmonic, as expected if neurons that can synchronize to both harmonics are responsible for detecting mistuning? Or is it possible that the timing patterns from neurons across the entire spectrum are displayed at some central processor and the "odd ball" identified at that level? To answer this question one can do mistuned harmonic discrimination experiments with diverse spectra. For instance, in the case of a mistuned 4th harmonic it is interesting to know what happens to performance if the 3rd and 5th harmonics are absent from the spectrum. The answer from experiment is that performance drops appreciably. One can reverse this idea and examine the effect of eliminating all the harmonics from the spectrum except for the 3rd, the 5th and the 4th, possibly mistuned. The answer is that performance is again adversely affected. The conclusion from these experiments and others is that the ability to segregate a mistuned harmonic appears to derive from a system that is tuned, but it is not tuned as narrowly as a critical band. For instance, some interaction with a second harmonic at 400 Hz aids the identification of a mistuned 4th harmonic near 800 Hz.

Level effects: Thus far the data from the mistuned harmonic experiment has led us to a model in which tuned neural elements identify an asynchronous signal, such as a mistuned harmonic, on the basis of a weak correlation with neighboring spectral components. Possibly these neural elements are part of the auditory periphery. But for a neuron to perform this function, it must synchronize to at least two spectral components. Peripheral neurons do show this kind of behavior, especially at low signal levels, near 40 dB. (Javel, 1980). At higher signal levels an eighth-nerve neuron increasingly tends to synchronize to only one of several components. This suggests that performance in identifying mistuned harmonics really ought to be best when the signal level is about 40 dB. Experiments with a mistuned 4th harmonic mistuned by 20 Hz and with a tone duration of 50 ms found a strong peak in performance in the region of 30 to 40 dB SPL for seven listeners out of seven.

SEQUENTIAL GROUPING

Sequential grouping has been studied with experiments on stream segregation, beginning with the work of Bregman and Campbell (1971) Of interest here is the conjecture that the ability to segregate auditory streams can be understood mainly in terms of peripheral channeling. Only two kinds of peripheral channel are allowed in this model, lateral and tonotopic.

This idea has been tested against experiments on interleaved melody identification, where the two interleaved melodies differed in a dozen different ways (Hartmann and Johnson, 1988). As expected, melody identification performance is excellent when the two interleaved melodies were in opposite ears or when they are played in different octaves. Performance is poor when the two melodies are played without such differences between them. Performance is good if one melody is played with a sine tone and the other with a complex tone, as expected because different peripheral channels are excited. Apparently one must think in terms of narrowly-tuned channels because performance is also good if one melody is played with odd harmonics and the other with a fundamental and even harmonics.

When tones of two melodies differ only in intensity or only in duration the peripheral auditory system does not provide much basis for segregation. Although some higher-level grouping process may segregate tones into different streams based on intensity or duration, there is little difference in peripheral channeling. Performance under these conditions is neither poor nor good. Similarly it is possible that higher-level processes mediate stream segregation in the case that tones differ in the shape of their amplitude envelopes or in their roughness. Experiments show, however, that melody identification is poor in these circumstances, presumably because tones that differ only in these ways still excited the same peripheral channels.

CONCLUSION

Auditory grouping processes are important in the perception of music. Psychoacoustical experiments suggest that much about these processes can be seen in the periphery of the auditory system, with peripheral neural elements behaving as expected in terms of tuning and synchronization. It is particularly remarkable that so much about sequential grouping, as observed in stream segregation experiments, can apparently be predicted on the basis of peripheral channeling. Of course, grouping models based on the periphery are incomplete. Higher-level processes that make decisions on the basis of the peripheral data are required, and these no doubt introduce important effects of their own. It is, however, interesting to observe the influence of expected peripheral operations in auditory grouping.

REFERENCES

Bregman, A.S. and Campbell, J. "Primary auditory stream segregation and the perception of order in rapid sequences of tones," J. Exp. Psych. **89**, 244-249 (1971).

Hartmann, W.M. "Perceptual entities from complex inharmonic tones," abst. Assn. for Research in Otolaryngology **8**, 168 (1985).

Hartmann, W.M. "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, G.M. Edelman, W.E. Gall, and W.M. Cowan eds. (John Wiley, New York, 1988).

Hartmann, W.M. and Johnson, D. "Stream segregation - source grouping vs peripheral channeling," J. Acoust. Soc. Am. abst. **84**, S143 (1988).

Javel, E. "Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones," J. Acoust. Soc. Am. **68**, 133-146 (1980).

McAdams, S. *Spectral fusion, spectral parsing and the formation of auditory images,* Ph. D. thesis, Stanford University (1984).

Moore, B.C.J. *An Introduction to the Psychology of Hearing*, (Academic, London, second edition, 1982).

Moore, B.C.J., Peters, R.W., and Glasberg, B.R. "Thresholds for the detection of inharmonicity in complex tones," J. Acoust. Soc. Am. **77**, 1861-1867 (1985).

Moore, B.C.J., Peters, R.W., and Glasberg, B.R. "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479-483 (1986).

Rasch, R.A. "The perception of simultaneous notes as in polyphonic music," Acustica **40**, 21-33 (1978)

Rasch, R.A. "Synchronization in performed ensemble music," Acustica **43**, 121-131 (1979).