Report 84: file: [tex] kyoto.tex, 28 Aug 1994

For the ATR Workshop
A Biological Framework for Speech Perception and Production
16-17 September 1994, Kyoto, Japan

ON THE PERCEPTUAL SEGREGATION OF STEADY-STATE TONES

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

ABSTRACT

Human listeners can perceptually segregate two different simultaneous tones, even if the spectra of the two tones are interleaved, and even if the two tones have coincident onsets. This remarkable ability can be profitably studied with psychoacoustical experiments using the mistuned harmonic paradigm, where the listener is required to detect a single mistuned harmonic in an otherwise periodic complex tone. A particular goal of the experiments is to decide whether the segregation process can be best explained as a spectral analysis or a temporal analysis.

Recent mistuned harmonic matching experiments have studied the dependence of segregation on mistuned harmonic number, amount of mistuning, fundamental frequency, tone duration, and level. The data support a temporal model of segregation in which the detection of neural asynchrony plays a preeminent role. The ability to match decreases with increasing mistuned harmonic number in a way that precisely parallels the loss of synchrony observed in physiological recordings from eighth-nerve neurons. Further, mistuned harmonic detection experiments show a nonmonotonic dependence on signal level that resembles the level dependence of multiple synchrony in the eighth nerve.

A question of current interest concerns the tuning of the decision process: whether synchrony/asynchrony is assessed within a narrow frequency channel or whether the process makes comparisons across the entire tonotopic axis. Experimentally, this question has been studied by various manipulations of harmonics that are near neighbors of the mistuned harmonic. Most of the data suggest that the system is tuned, though the channels appear to be broader than typical critical bands.

INTRODUCTION

When we look at two separated objects, images are formed at separated places on the retina where the neural processing of visual stimuli begins. If the angular separation of the objects is great enough the visual system can resolve the separate images, and it registers the fact that there are two objects out there, or anyhow that there are more than one.

When we listen to two distinct sine tones, patterns are formed at separated places on the basilar membrane where the neural processing of auditory stimuli begins. If the frequency separation between the sine tones is great enough (a little less than a critical bandwidth) the auditory system can resolve the separate tones, and it registers the fact that there are two tones.

When we look at two objects that overlap or for which the retinal images are interleaved, the initial neural stimulus is not an adequate basis for identifying, or even counting, the objects. The identification of overlapping objects requires pattern recognition processes with intricate rules of inference (Marr, 1982).

When we listen to two complex tones with interlaced partials the spatial pattern on the basilar membrane (or any other tonotopic coordinate) is not an adequate basis for determining the pitch or tone color of either tone. The pattern recognition process by which individual auditory entities are extracted from a collection of interlaced partials is known as segregation and integration (Hartmann, 1988).

The partials of a periodic complex tone are not normally heard out individually. They fuse perceptually to create a single entity characterized by a pitch and tone color. This is the process of integration. When there are two complex tones, not all the the partials are integrated into one entity. The auditory system appears to sort the partials in a way that creates several different entities. This is the process of segregation.

In listening to speech, music and the sounds of the everyday environment, the most important basis for the segregation of different entities is in the onset. Partials generated by a single source tend to start together, and their starting time is generally slightly different from the starting time of the partials of a different source. It is tempting to suppose that the superb temporal resolution of which the auditory system is capable actually evolved in order to perform just such processes of segregation. The psychophysical effects of asynchronous onsets, especially as they occur in the perception of polyphonic music, have been studied by Rasch (1978, 1979). A related study by McAdams (1989) has shown how common frequency modulation can lead to integration while dissimilar modulation can promote segregation.

A further basis for segregation and integration results from the tendency of many important sounds to be periodic and therefore to have harmonic partials. This is true of the sounds of sustained-tone musical instruments and of the vowels of human speech. The periodicity appears to be a basis for integration that applies to the steady-state part of a sound, independent of onsets.

The significance of periodicity in integrating the harmonics of a single tone and the significance of different periodicities in segregating different tones immediately raises questions about the right kind of model to use in thinking about the process. Should the model emphasize the periodicity, as it appears in the time domain, or should the model emphasize the harmonic structure in the spectral domain?

Segregation and integration are essentially matters of pitch perception because tones, either one or more, are identified primarily by their pitches. Therefore, it is natural that issues of timing and spectrum that arise in pitch perception also appear in segregation and integration.

The residue theory of pitch perception is a timing model that is specifically designed to deal with the sense of low pitch that one perceives in a periodic waveform. As a model of the fundamental pitches of complex tones, the residue theory is now discredited. However, there are timing models for the pitch of a pure tone which are still useful, especially in explaining the pitches of short tones (Goldstein and Srulovicz, 1977). Timing models of pitch have gained particular credibility recently from studies of pitch perception in individuals with cochlear implants.

A timing model for integration and segregation emphasizes the idea that the various harmonics of a complex tone are phase locked and therefore the neural spikes in different frequency channels are synchronized. It is the synchrony that leads to integration of the channels. Correspondingly, a failure to find an appropriate synchrony in all tuned channels is evidence that leads to the perception of two or more tones. It is evident that segregation and integration can be based upon synchrony or deviations from synchrony only in those tuned channels where the neurons can encode synchrony. This limits the use of synchrony to low frequencies.

The spectral model of segregation and integration is a theory that is consistent with modern approaches to complex tone pitch perception. It focuses on spectral components that are resolved by the auditory system and subjects them to a template fitting process. Different aspects of template fitting are emphasized in pitch models by Goldstein (1973) and Terhardt (1974). The integration/segregation model of Duifhuis, Willems and Sluyter (DWS, 1982) regards the template as a sieve. Spectral components that pass through the sieve are integrated into a complex tone and contribute to the pitch and timbre of the tone. Components that do not pass are segregated into some other percept. In favor of this model is the fact that it is able to segregate two simultaneous speech sounds with the same success as human listeners.

THE MISTUNED HARMONIC EXPERIMENT

To answer questions about timing and/or spectrum as bases for integration and segregation we have studied the detection of mistuned harmonics. Consider a complex periodic tone with a fundamental frequency near 200 Hz and a number of strong harmonics. Such a tone is perceived as a single entity with a low pitch and a bright or buzzy timbre.

If one of the harmonics of the complex tone is slightly mistuned from its correct harmonic frequency, several effects can occur. First, there is a change in pitch of the complex as a whole (Moore *et al.*, 1985). This happens because the auditory system computes the low pitch of a complex tone as a weighted average of frequency information from the various partials of the tone. Second, there are "beats of mistuned consonances" (Plomp, 1967), reflecting a sensitivity in the auditory system to dynamic phase changes. Most importantly, the mistuned harmonic may be heard out from the complex as a whole as a separate entity. If, for example, the fourth harmonic is mistuned, the listener may become aware of a flute-like tone playing the double octave and accompanying the buzzy tone having the low pitch. It is this latter effect that interest us.

Two psychoacoustical paradigms have been used to study the detection of the mistuned harmonic. One of them is the *mistuned harmonic matching experiment.* Here the listener's task is to match the pitch of the mistuned harmonic in a complex tone by adjusting the frequency of a sine tone. The complex tone and the sine tone are presented successively, not simultaneously. The listener does not know which harmonic is mistuned on a given trial; it might be any one from the fundamental to the 16th harmonic. If the listener matches the mistuned harmonic correctly, he scores a "hit". Otherwise the match is called a "miss." What is particularly attractive about this paradigm is that if the listener can successfully perform a pitch match then we know that he has heard out the mistuned harmonic as a separate entity. It is impossible to make a successful match based upon the other effects of a mistuned harmonic, the shift of the low pitch or the beats of mistuned consonances.

A more efficient paradigm is the *mistuned harmonic discrimination experiment.* Here the listener hears two tones in succession. In one tone the partials are all perfect harmonics. In the other, randomly the first or the second, there is a mistuned harmonic. The listener's task is to identify the tone that includes the mistuned harmonic. Because the task is not specific, one must guard against the artifacts caused by pitch shift cues and beats of mistuned consonances (Moore, *et al.* 1986). To minimize the role of beats, experiments using the discrimination paradigm use tones with duration less than 100 ms, usually less than 50 ms.

Matching Experiments and Synchrony

Mistuned harmonic matching experiments give evidence that neural synchrony plays an important role in segregating a mistuned partial in a complex tone (Hartmann, *et al.*, 1990). In these experiments the mistuned harmonic number and the fundamental frequency were systematically varied. The data showed that the most important determinant of the listeners ability to segregate is the frequency of the mistuned component. Unlike the predictions of models based upon spectral resolution, and unlike the DWS model, segregation is not a simple function of mistuned harmonic number. Instead, the ability to segregate drops dramatically with increasing frequency, between 2 and 3 kHz. This effect is so dramatic that it even dominates the effect of changing the amount of mistuning, which is the major controlled variable of the experiment. Comparison with synchrony measurements made on auditory neurons in cat shows that the drop in performance on the segregation task precisely parallels the decrease in maximum possible synchrony as a function of frequency. This supports the timing model of segregation, because it is only at frequencies where synchrony is possible that the modulated synchrony, or asynchrony, associated with a mistuned partial can be noticed.

Discrimination Experiments and Autocorrelation

Having identified an important role for synchrony, or neural timing, in the segregation of mistuned harmonics, it is natural to wonder how timing is used in the process. There is information on this question in the results of mistuned harmonic discrimination experiments. These are parametric studies, and there are a lot of parameters to vary: fundamental frequency, mistuned harmonic number, amount of mistuning, duration of the tones, relative phases of the partials, and signal level. As it turns out, all of them matter.

Level effects

Mistuned harmonic detection experiments using mistuned 4th, 5th and 6th harmonics of a 200 Hz complex tone find that the ability to detect a mistuned harmonic is a non-monotonic function of signal level. The function has a maximum at a level of about 40 dB SPL. A possible explanation for this maximum can be found in the neural synchrony studies of Javel (1980) and Greenberg *et al.* (1983), where it was found that a neuron of the eighth nerve can synchronize to both of two different frequencies if the level is about 40 dB. At higher levels one of the two, usually the tone of lower frequency, dominates the synchrony.

Amount of Mistuning

There is one parameter whose effect seems *a priori* obvious, namely, the amount of mistuning of the mistuned harmonics. When the mistuning is zero there is no basis for discrimination - the two tones are identical. One expects that the greater the mistuning, the more distinguishable the tones should become and the higher the percentages of correct responses in the forced-choice task should be. The experimental test of this idea used a 200 Hz tone with seven harmonics of equal amplitude at an overall level of 40 dB SPL. The duration of the tones was 50 ms. The fourth harmonic was mistuned, and the amount of mistuning was a parameter. As expected, increased mistuning led to improved detection. However, the detectability showed a plateau at a mistuning of 20 Hz, corresponding to one synchrony cycle ($20 \text{ Hz} \times 50 \text{ ms} = 1$). Similarly, when the duration of the tones was reduced to 30 ms, the detection plateau occurred near 33 Hz. Autocorrelator models, such as the tuned autocorrelator model (Hartmann, 1986) or the summary autocorrelogram of Meddis and Hewitt, (1991 a,b) are capable of predicting the plateau at one synchrony cycle.

The Tone Duration

A prediction of the autocorrelator models is that if one does an experiment with the amount of mistuning held constant and the tone duration varied as a parameter, some oscillatory behavior might appear, corresponding to synchrony cycles. Apart from this autocorrelation model, there is no reason to expect anything other than a monotonic improvement in performance with increasing tone duration.

The experiments to test the role of tone duration were identical to the experiments above except that the amount of mistuning of the fourth harmonic was fixed at 20 Hz, and the tone duration was varied. On the basis of the model one expects a plateau in the region of 50 ms, and that is just what the detectability data showed. In some cases, the data actually had a local maximum at the duration of a synchrony cycle.

EVIDENCE FOR TUNED CHANNELS

If it is accepted that mistuned harmonics are segregated from the background on the basis of synchrony anomalies, it remains to discover the locus of those anomalies, whether they occur in tuned channels or at a site that looks across all tuned neural channels. We have studied the question of tuning with two kinds of mistuned harmonic experiments, gap experiments and interference experiments. Both of them are of the discrimination type.

The gap experiments test the idea that a mistuned harmonic is segregated because it fails to exhibit a common synchrony with its neighbors in the same tuned channel. If the neighbors are removed there would no longer be any basis for judging the synchrony within the channel. Gap experiments were done with a mistuned fourth harmonic in which the third harmonic, or the fifth harmonic, or both were missing. The data showed that missing neighbors, especially a missing third harmonic, caused a marked decrease in the detectability of the mistuned fourth.

The interference experiments attempt to interfere with the detection of a synchrony anomaly, caused by a mistuned target harmonic, by some other mistuning, which also leads to a synchrony anomaly. Specifically, the task was to detect a mistuned fourth harmonic, mistuned by a small amount on one of the two intervals, in the presence of a mistuned second harmonic, mistuned by a large amount on both intervals. The data showed a small decrement in detectability of the mistuned fourth caused by the interference of the mistuned second. Considerably greater decrement occurred when the interfering mistuned harmonic was changed to the sixth. By contrast, interference from a mistuned eighth was negligible. Thus, interference exhibits tuning, suggesting that synchrony is evaluated in tuned channels, indicating auditory filtering. Both the gap experiments and the interference experiments suggest a filter with a high frequency slope that is considerably steeper than the low frequency slope.

CONCLUDING DISCUSSION

The perceptual operations of integration and segregation can be approached in terms of both tonotopic spectral template fitting models and neural synchrony timing models. The mistuned harmonic experiment gives considerable evidence to support the idea that listeners segregate mistuned harmonics on the basis of anomalies in neural synchrony. The rapid decline in the ability to detect a mistuned harmonic with increasing frequency parallels the loss of synchrony in the mammalian ear as observed in physiological studies. The nonmonotonic dependence of detectability on signal level parallels the dependence of multiple synchronies observed physiologically. Structure in the dependence of detectability upon the amount of mistuning and the duration of the stimulus can both be understood from an autocorrelation model for synchronous neural spikes. By contrast, the tonotopic template fitting models do not account for the above effects. To the extent that the mistuned harmonic experiment is a representative segregation operation, neural synchrony is an important part of the integration and segregation of the complex tones in music and speech.

Mistuned harmonic detection experiments employing spectral gaps or interfering tones suggest that neural synchrony is evaluated in tuned channels. This means that synchrony anomalies caused by a mistuned harmonic are recognized by comparison with only a small set of neighboring harmonics. The experiments do not suggest that there is an overseer that scans across the entire tonotopic axis. Ultimately, of course, there must be some process that combines the information from different tuned channels to form the integrated entities that we recognize as music and speech. The mistuned harmonic experiments, however, indicate that there is preliminary synchrony evaluator that is tuned. On the other hand, a long series of experiments on the pitches of mistuned harmonics, none of which was discussed in this paper, appear to be suggesting that mistuned harmonics are actually perceived with respect to the entire background of the other harmonics in the complex. In fact, the most straightforward explanation for the recent pitch results lies in a model that resembles a template fitting model (Lin and Hartmann, 1994). The resolution of the detection data with the recent pitch data must await another time, another workshop, and another paper.

ACKNOWLEDGEMENT

This work was supported by the United States National Institutes of Health.

REFERENCES

Duifhuis, H., Willems, L.F., and Sluyter,R.J. (1982) "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am., **71**, 1568-1580.

Goldstein, J.L. (1973) "An optimal processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496-1516.

Goldstein, J.L. and Srulovicz, P. (1977) "Auditory-nerve spike intervals as an adequate basis for aural spectrum analysis," in *Psychophysics and Physiology of Hearing*, ed. E.F. Evans and J.P. Wilson, (Academic, New York) pp.337-345.

Greenberg, S., Geisler, C. D., and Deng, L. (1983) "Filter characteristics of low-frequency cochlear nerve fibers as determined by synchrony response patterns to two-component signals." J. Acoust. Soc. Am. (abst) **74**, S7.

Hartmann, W.M. (1986) "Pitch and the perceptual organization of complex spectra," J. Acoust. Soc. Am. (abst.) **79**, S65.

Hartmann, W.M. (1988) "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, ed. G.M. Edelman, W.E. Gall and W.M. Cowan (Wiley, New York) pp.623-645.

Hartmann, W.M., McAdams, S., and Smith, B.K. (1990) "Matching the pitch of a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712-1724.

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

Johnson, D.H. (1980) "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115-1122.

Lin, Jian-Yu and Hartmann, W.M. (1994) "The pitches of mistuned harmonics," J. Acoust. Soc. Am. (abst) **95**, pp 2965.

Marr, D. (1982) *Vision*, W.H. Freeman, San Francisco.

McAdams, S. (1989) "Segregation of concurrent sounds. I: Effects of frequency modulation coherence." J. Acoust. Soc. Am. **86**, 2148-2159.

Meddis, R. and Hewitt, M.J. (1991a) "Virtual pitch and phase sensitivity of a computer model of the auditory periphery I." J. Acoust. Soc. Am. **89**, 2866-2882.

Meddis, R. and Hewitt, M.J. (1991b) "Virtual pitch and phase sensitivity of a computer model of the auditory periphery II." J. Acoust. Soc. Am. **89**, 2883-2894.

Moore, B.C.J., Glasberg, B.R. and Peters, R.W. (1985) "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853-1859.

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

Plomp, R. (1967) "Beats of mistuned consonances," J. Acoust. Soc. Am. **42**, 462-474.

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

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

Terhardt, E. (1974) "Pitch, consonance and harmony," J. Acoust. Soc. Am. **55**, 1061-1069.