The role of FM-induced AM in dynamic spectral profile analysis

Stephen McAdams & Xavier Rodet

Institut de Recherche et Coordination Acoustique/Musique (IRCAM) 31, rue Saint-Merri F-75004 Paris, France

Introduction

Recent work by Green and colleagues (cf. Green, 1983) has begun to demonstrate with psychophysical methods that the human auditory system is capable of extracting a global representation of the spectral envelope of a signal. This may serve to identify the resonance structure characteristics of a sound source. However, this previous work has been exclusively confined to relatively dense steady-state, inharmonic stimuli of simple spectral forms (a single bump in the spectral profile). Dynamic stimuli, those with frequency jitter or vibrato, might be helpful in reducing perceptual ambiguity in cases where there are not enough partials present in a sound to clearly define a spectral envelope. They may do this by tracing out the spectral envelope through time, thus increasing information about the resonance structure.

While this seems intuitively obvious, previous work on high-pitched vowel identification in the presence of frequency vibrato has yielded ambiguous results (Sundberg, 1977). These researchers claim that intonation contours or vibrato had slight, or even detrimental, effects on vowel identification. The present study demonstrates, to the contrary, that if the amplitude behavior of a given partial is coupled with its frequency behavior according to a given spectral envelope, this information can be used by the auditory system to discriminate and identify the vowel quality or timbre of complex harmonic sounds.

Experiment 1: Spectral envelope discrimination

Stimuli

The experiment was conducted with both harmonic complex and sinusoidal stimuli. All tones had a duration of 1 sec and an amplitude envelope with raised cosine attacks (150 ms) and decays (200 ms). Complex tones consisted of the first 8 harmonics of a 675 Hz fundamental. Sine tones were at 1350 Hz, equivalent to the second harmonic of the complex. This high F_0 gives wide harmonic spacing and thus a poor spectral envelope definition in the absence

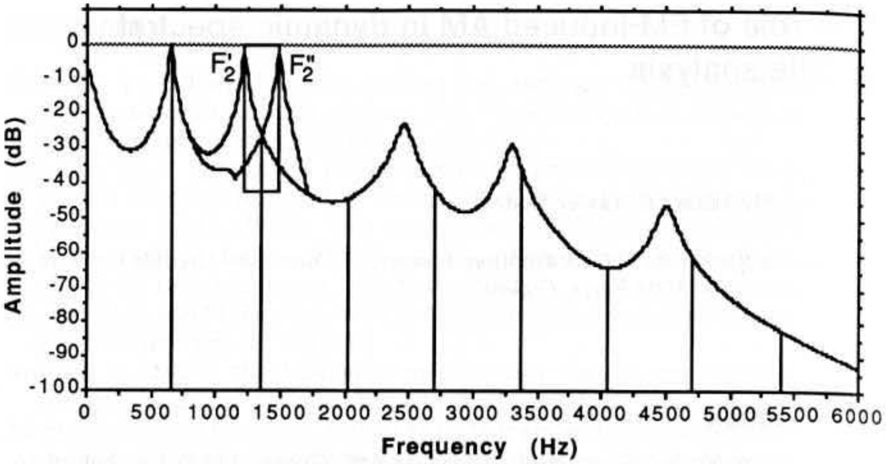

Figure 1. The two spectral envelopes used in the experiment.

of frequency modulation.

Eight peak-to-peak widths were used throughout: 0, 1, 2, 3, 4, 6, 8, 10% of the partials' frequencies. A psychometric function was determined based on this dimension. The vibrato was a sinusoidal frequency modulation with a frequency of 6.5 Hz. The starting phase of the vibrato was randomly selected from 0, $\pi/2$, π, $3\pi/2$. This was necessary to avoid discrimination judgments based on counting the number of peaks in the amplitude envelope for the modulated pure tone condition, since with sine phase vibrato the amplitude would always increase at the beginning for one spectral envelope, and decrease for the other.

The two table-lookup spectral envelopes applied to the components, SE1 & SE2, are illustrated in Figure 1. They are like allophones of /ʌ/ with SE1 being closer to /o/ and SE2 being closer to /æ/. The envelopes were stored in a table which returned the instantaneous amplitude corresponding to the instantaneous frequency of each partial. In this way, the vibrato was coupled to an amplitude modulation defined by the spectral envelope function. The spectral envelopes consisted of 5 formants, 4 of which were identical in the two cases. The only difference was the center frequency of the second formant (F'_2=1215 Hz for SE1, F''_2=1485 Hz for SE2). The skirts of these 2 formants intersect at the center frequency of the second harmonic, and the main difference between them in the region just around this harmonic (at small vibrato widths) is a change in the sign of the spectral slope. The part of the slopes in the boxed area in Fig. 1 covers the range for a 20% peak-to-peak vibrato, and was constructed over this range such that the two envelopes are mirror images (in linear frequency) of one another about the 2nd harmonic frequency. The envelopes were also constructed so that the frequency-amplitude coupling for all other harmonics was identical in the 2 cases.

In the non-roving global amplitude condition, complex tones were presented at 75 dBA and sine tones were presented at 58 dBA, this latter being equal to the intensity of that tone within the complex. The rms amplitudes of the stimuli at a given vibrato width (2 spectral envelopes at 4 vibrato starting phases) were identical. In the roving amplitude condition the tones were presented at the above intensities or at that value ± 5 dB.

Method

Stimuli were presented diotically over headphones. Each trial consisted of a sequence of four tones arranged in two pairs, with each pair constituting one observation interval. In one interval, both tones had the same spectral envelope (SE). In the other interval, they had different spectral envelopes. Four trial structures are thus possible: (SE1 SE1 / SE1 SE2), (SE1 SE2 / SE1 SE1), (SE2 SE2 / SE2 SE1), and (SE2 SE1 / SE2 SE2). These structures were counterbalanced across trials. The starting phase of the vibrato was randomly selected for each of the 4 tones. In the roving condition the amplitude of each tone was randomly selected from the 3 possible values.

The subject's task was to identify the interval containing the "different" pair by pressing a corresponding button. Feedback indicating the correct response was given.

Trials were presented in blocks of 80 with ten repetitions of each of the 8 vibrato widths in random order.

Subjects completed several training blocks until their psychometric functions appeared to stabilize. The number of training blocks varied considerably between subjects (5-35). Then ten blocks were collected, giving a total of 100 2IFC judgments for a data point at each vibrato width. The % correct measures at each vibrato width were averaged across the 10 blocks in order to obtain the mean and standard deviation. This latter statistic was then used to evaluate the reliability of the 75% threshold on the psychometric function as described in the next section.

Four subjects completed the non-roving conditions for complex and sine tones in that order. Afterward, two of these subjects completed the roving conditions for complex and sine tones.

Results

A spline curve was fitted to the 8 data points for each subject in each condition and the 75% point was determined as a measure of threshold performance. In order to have a measure of the variability of this threshold, spline curves were also fitted to points 1 standard deviation above and below the means from which the 75% points were also determined. One notes that these outer points are asymmetric with respect to the mean. Therefore, as a rough measure of the overall variation in the threshold the mean distance between the center point and the outer 75% points was calculated. This measure was used as an estimate of the standard deviation along the stimulus dimension. It is probably an over-estimation of the standard deviation, making the statistical test very conservative.

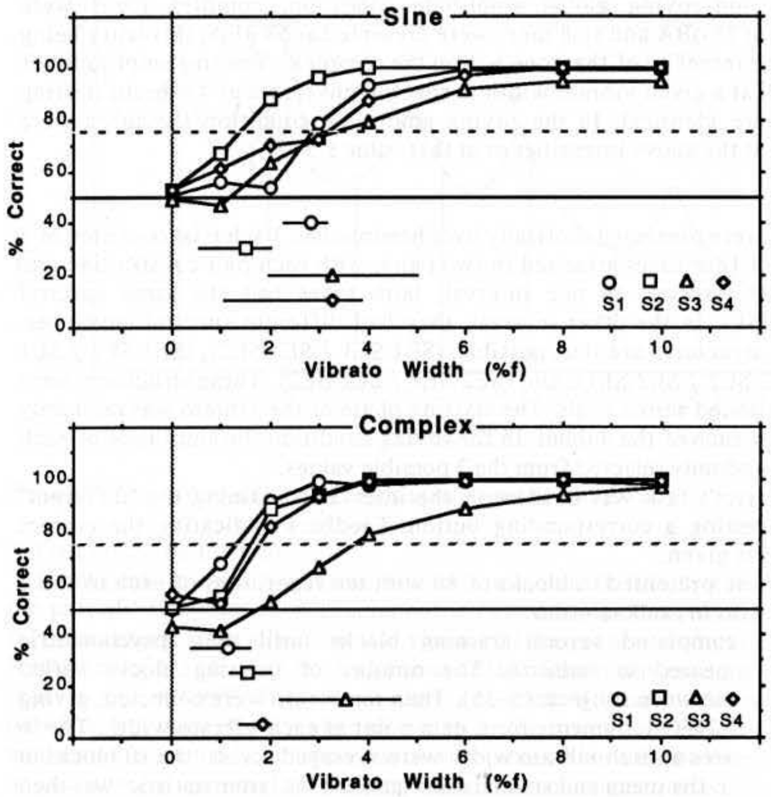

Figure 2. Psychometric functions for 4 subjects (non-roving condition) showing % correct discrimination as a function of vibrato width in %f. The points with horizontal bars in the lower portion of the plot indicate the 75% point of the mean curve and the estimated range of standard error for each subject.

The mean data for 4 subjects are shown for the non-roving condition for both complex and sine tones in Fig. 2. All Ss attain near-perfect performance in the range of vibrato widths used indicating that the stimulus difference is easily discriminable. All psychometric functions are monotone increasing indicating that the perceptual factor upon which discrimination is based varies with vibrato width. The range of threshold vibrato widths for the 4 Ss was 1.2-3.8% (16-51 Hz at the 2nd harmonic). For S1 & S4, complex thresholds are significantly lower than sine thresholds (p<.01; t-test). They are approximately equal for S2 & S3. This may indicate differences in decision strategies among the Ss. If complex thresholds were always less than sine thresholds, one might hypothesize that the same dynamic frequency-amplitude slope information was more easily interpreted in the global context given by the behavior of the other harmonics. Such may be the case for S1 & S4.

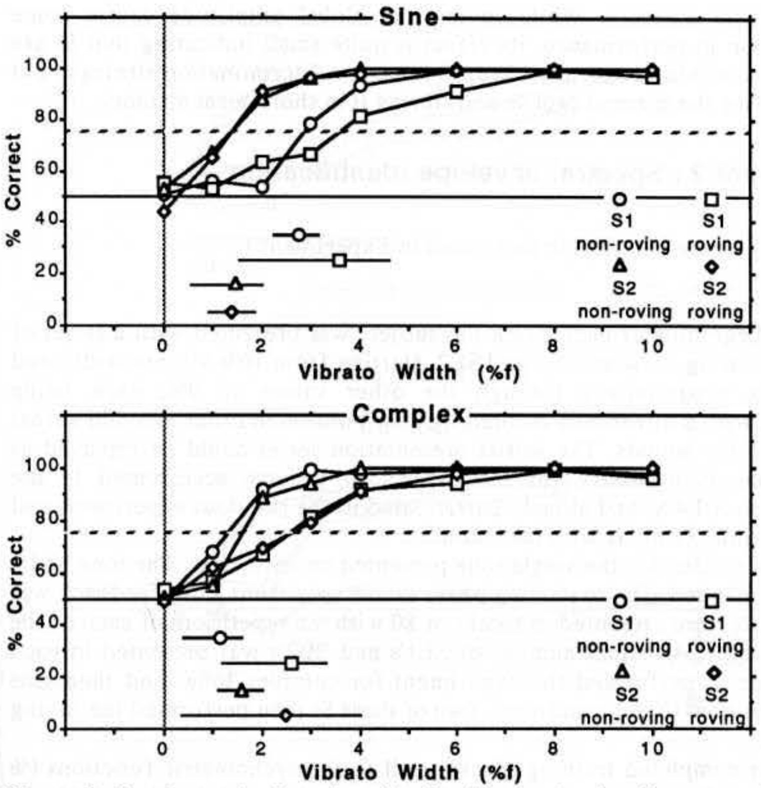

Figure 3. Psychometric functions for 2 subjects showing % correct discrimination as a function of vibrato width. Points with horizontal bars as in Fig. 2.

The mean data comparing roving and non-roving conditions for Ss 1 & 2 are shown in Figure 3. Both subjects attain near-perfect performance by 6% peak vibrato width in both of these conditions. For complex tones, the thresholds for roving amplitude stimuli appear to be greater than those without roving for both Ss. Only the difference for S1 is statistically significant (p<.01), though it is relatively small. For sine tones, neither S shows an effect of amplitude roving. Comparing sine with complex tones within the roving condition, S1 shows no difference while S2 does, complex tones having a higher threshold (p>.05).

We might conclude from these data that the difference in spectral envelope following for minimally different envelopes is indeed possible at relatively low vibrato widths (1.2-3.8%). This discrimination is easier for some Ss in the presence of a vowel-like spectral envelope on flanking harmonics than it is with a single sinusoid. While roving the global amplitude causes some deterioration in performance, its effect is quite small indicating that Ss are not using a within-frequency channel intensity discrimination strategy, but are extracting the spectral profile and storing it in short-term memory.

Experiment 2 : Spectral envelope identification

Stimuli

The stimuli are identical to those used in Experiment 1.

Method

At the beginning of each block the subject was presented with a series of tones alternating between SE1 and SE2, starting from 10% vibrato width and descending progressively through the other values to 0%, each being associated with a differently colored light and button in order to avoid verbal labeling of the sounds. The initial presentation series could be replayed as many times as necessary for the subject to become accustomed to the differences. All 4 Ss had already participated in the previous experiment and were thus quite familiar with the stimuli.

Ss were to identify the single tone presented on each trial. The tone had a randomly selected vibrato starting phase as in Experiment 1. No feedback was given. Trials were presented in blocks of 80 with ten repetitions of each of the vibrato widths. An equal number of SE1's and SE2's was presented in each block. Four Ss performed the experiment for complex tones and then sine tones in the non-roving condition. Two of these Ss then performed the roving condition.

Subjects completed training blocks until their psychometric functions (% correct identification as a function of vibrato width) stabilized. Ten blocks were then collected from which were determined the mean % correct and standard deviation across the 10 blocks. From these values the 75% threshold and range of its standard error were calculated from fitted spline curves as in Experiment 1.

Results

The mean data for 4 Ss are plotted for the non-roving condition for both types of tone in Figure 4. Three of the Ss attain near-perfect performance at a vibrato width of 8% for both sines and complex tones. The range of threshold vibrato widths is 0.6-3.6% (8-49 Hz at the 2nd harmonic). S3 attains threshold for sine tones at 6.2%, but never achieves better than chance performance for complex tones even at the largest vibrato width. This subject claimed to have tried several criteria for identification, but could not latch onto anything that allowed positive identification. With that one exception, all curves are monotone increasing indicating that the perceptual factor permitting identification varies with vibrato width.

Reversing the trend in Experiment 1, Ss 2 & 3 have different thresholds for complex and sines in this identification task (p<.05 for S2; threshold for complex well beyond stimulus range for S3), while Ss 1 & 4 show no difference. The difference shown by S2 is from 0.6% for sines compared to 1.2% for complex tones. The threshold for sines may well be lower since more data would be needed between 0 and 1% to accurately estimate the threshold.

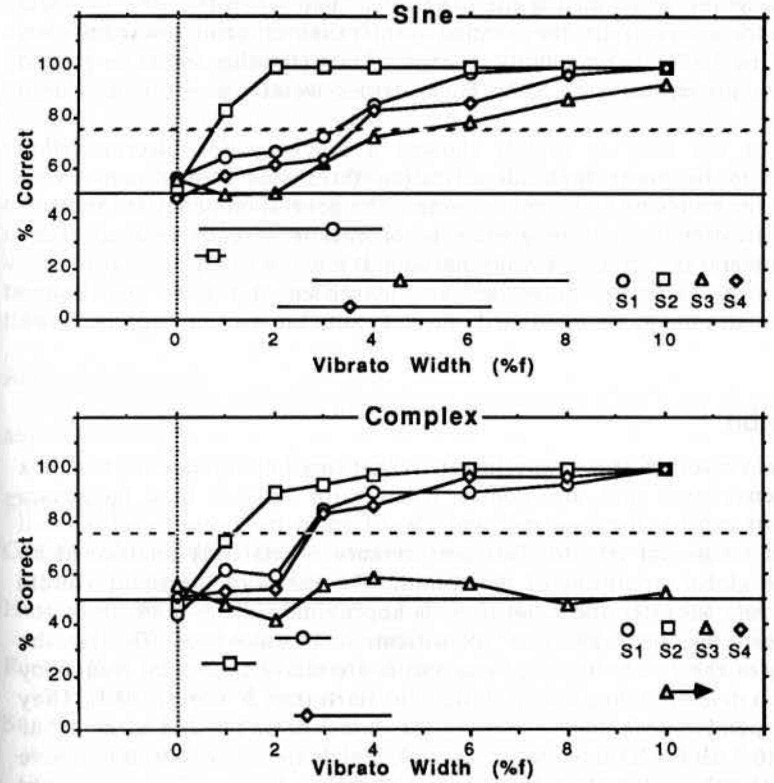

Figure 4. Psychometric functions for identification (non-roving condition). Points with horizontal bars as in Fig. 2.

The psychometric functions for roving amplitudes (Ss 1 & 2) showed no differences between sine and complex tones nor between roving and non-roving conditions.

These data suggest that even very similar spectral envelopes can be successfully identified in the presence of vibrato (and with a severely restricted set of choices), when the fundamental frequency is fairly high and the spectral envelope is not well defined by the relative amplitudes of the frequency components in the absence of vibrato.

A possible criticism of this experiment may be that the stimulus set was too small, making the "identification" experiment one of "discrimination across trials". All Ss remarked during the experiment that when the vibrato width was small for several trials, they tended to shift their criterion toward the less bright of the 2 SE's. Subsequently, a large vibrato stimulus would be judged (sometimes erroneously) as SE2. These errors would raise the measured thresholds.

Three of the four Ss (1,3,4) showed a tendency for discrimination thresholds to be lower than identification thresholds in the non-roving condition for both sine and complex tones. The general tendency seems to be for identification to require greater perceptual difference than discrimination. It seems intuitively obvious that sounds must be easily distinguishable in order to be correctly categorized and identified, but given the limited number of stimuli to be identified, these results must be interpreted with caution.

Discussion

The main result of these experiments is that small differences in complex spectral envelopes that are poorly filled with only a few frequency components can be discriminated and identified in the presence of a small amount of sinusoidal vibrato. This performance is relatively unaffected by roving the global amplitude of the stimuli. The range of threshold vibrato widths across subjects and conditions is approximately 0.6-3.8% peak-to-peak (excepting the 6.2% sine identification threshold of S3). At the frequency of the 2nd harmonic, these values are above sinusoidal frequency modulation detection threshold (0.1-0.3% in Hartmann & Klein, 1980). They also correspond to peak-to-peak amplitude variations on the 2nd harmonic of 0.6-4 dB (6.7 dB for S3 sine identification), which are also either at or above detection threshold (0.8 dB at 1000 Hz and 60 dB; cf. Riesz, 1928). One might conclude that the modulation must be detectable along both dimensions in order for its combined effect to be extracted as a spectral envelope tracing.

The apparently small difference in performance introduced by the presence of other harmonics around the 1350 Hz 2nd harmonic (whose frequency-amplitude coupling defines other regions of the spectral envelope) bears some consideration. For the 2 Ss who show this trend (Ss 1 & 4), the average increase in threshold when the flanking harmonics are removed is only on the order of 1% of the component frequency. This corresponds to an average increase of about 1 dB in the peak-to-peak amplitude fluctuation on the 2nd harmonic. What this may suggest is that the information already present in the modulated sine tone is sufficient to explain performance with the complex tones. The reports from Ss about what they listened to in order to make the judgments on sine tones varied. Ss 1 & 4 felt that they were listening for a tone color difference in the two sines. Ss 2 & 3 felt that they were listening to a pitch difference. This latter criterion is easily understood since the amplitude of the sine is greater at lower frequencies for SE1 and greater at higher frequencies for SE2. A strategy that consisted of accumulating a weighted pitch representation of the tone and deciding whether it was the higher or lower would be successful since the discrimination thresholds are above frequency discrimination threshold if one were to measure the distance between the lower excursion of SE1 and the upper excursion of SE2. All Ss, however claimed to use tone color or vowel quality as the cue with the complex tones. It is entirely possible that Ss 2 & 3 used different strategies in the two cases.

Whatever the mechanism responsible for this performance, it is clear that one must take into account the dynamic nature of the stimuli. The basilar membrane activity pattern proposed by Green (1983) as the stimulus structure used to make the profile comparison is never present at any given moment in these stimuli. It is thus necessary for the auditory system to accumulate it through time.

Acknowledgments

The idea for this research arose from discussions with William Hartmann and Yves Potard.

References

Green, D.M. (**1983**). "Profile analysis: A different view of auditory intensity discrimination," American Psychologist **38**, 133-142.

Hartmann, W.M. & Klein, M.A. (**1980**). "Theory of frequency modulation detection for low modulation frequencies," J.Acous.Soc.Am. **67**, 935-946.

Riesz, R.R. (**1928**). "Differential sensitivity of the ear for pure tones," Phys.Rev. **31**, 867-875.

Sundberg, J. (**1977**). "Vibrato and vowel identification," Archives of Acoustics, Polish Academy of Sciences **2**(2), 257-266.

Comments

Moore:

You suggest that your results indicate that subjects do not use a within-channel strategy, but rather extract the spectral profile and detect changes in that profile. However, it would be possible for the subjects to perform the task using a within-channel strategy. Consider the behaviour of the second harmonic, at 1350 Hz. For one stimulus, increases in frequency are coupled with increases in amplitude, while for the other increases in frequency are coupled with decreases in amplitude. If the subject were to listen to the output of an auditory filter centred somewhat below 1350 Hz, say at 1100 Hz, then the depth of modulation at the output of that filter would differ for the two stimuli; it would be greater for the second than for the first. Thus the depth of modulation within a channel would provide a cue.

Reply by McAdams and Rodet:

This is a reasonable criticism which reflects more on the nature of the experimental task (2 IFC with feedback - which allows the development of such listening strategies) than on real-world behavior. If we assume a slope of 27 dB/Bark on the lower side of the activity pattern resulting from stimulation at 1350 Hz (Zwicker and Feldkeller, 1967), the amplitude modification depth of the activity in a frequency channel at 1100 Hz may be estimated for SE 1 as approximately four times that for SE 2 at any of the vibrates width used in this experiment (see Fig.C1). Since the slope is independent of the sound pressure level of the component, the difference in modulation depth between SE 1 and SE 2 is independent of the global amplitude and would thus be unaffected by amplitude roving. For this cue to be usable, the absolute level of activity in this channel needs to be detectable. Based on the above slope estimate and a 58 dB level of the component, the level of stimulation at 1100 Hz varies around a value of 21.5 dB. It is also necessary that the modulation depth be sufficient to be detectable. The estimated depths for stimuli with vibrato widths around threshold discrimination are listed below:

	SE 1	SE 2
4%	10.9 dB	2.8 dB
2%	5.4 dB	1.4 dB
1%	2.7 dB	0.7 dB

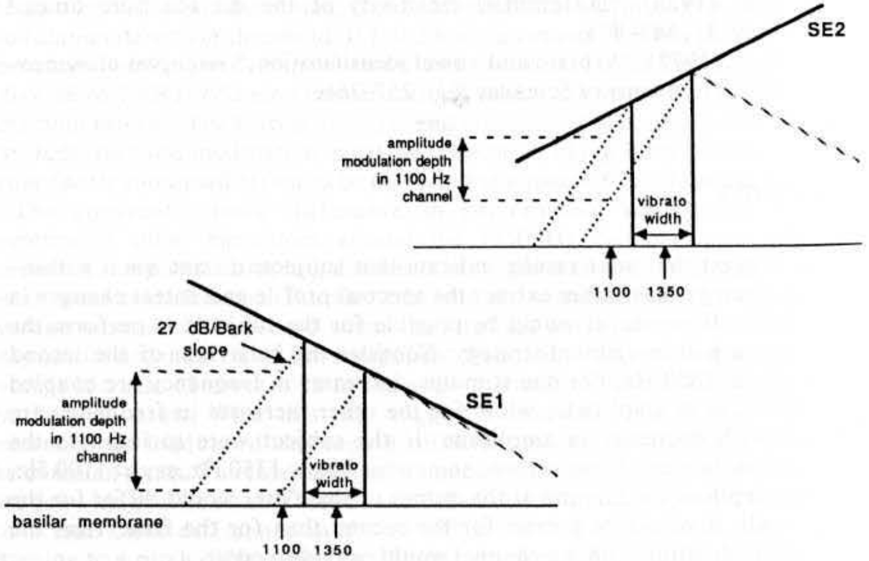
SE2
amplitude modulation depth in 1100 Hz channel
vibrato width
1100
1350
27 dB/Bark slope
SE1
amplitude modulation depth in 1100 Hz channel
vibrato width
basilar membrane
1100
1350

Figure C1. Frequency modulation of the basilar membrane activity pattern due to the component at 1350 Hz yields amplitude modulation depth in a channel at 1100 Hz that are larger for SE1 than for SE2.

If a listener attends to this channel, the amplitude modulation should always be detectable, at least for SE 1. However, no subject reached threshold discrimination at 1%. Also, if this cue was responsible for the discrimination observed, there should be no difference in performance between sine and complex tone conditions. Such a difference does exist for 2 Ss, though it is small. One of the subjects for whom there is no difference has very high thresholds, perhaps indicating that this information is not used. In any case, for these spectral envelopes, even roving the F_0 ± 5% across tones in a trial would not succeed in completely thwarting the possibility of using this kind of listening strategy. Different envelopes would need to be chosen to allow a greater frequency roving range.

Houtsma:

As shown in your figure you have chosen your stimuli such that for zero % vibrato width the stimuli were identical and, consequently, indiscriminable. Could you tell what would happen if for zero% modulation the second harmonics of the alternative signals are made unequal? This would allow us to assess the relative importance of frequency modulation compared with fixed differences between partials in the complex-tone discrimination task.

Reply by McAdams:

That would, of course, be more like the steady-state profile analysis experiments with wide component spacing and a small number of components, but with harmonic frequencies and a multi-formant profile. We have planned some studies to compare modulated and steady-state thresholds in spectral envelope discrimination and to compare performance on harmonic, multi-formant stimuli with the inharmonic flat spectrum used by Green and colleagues.

The **8th International Symposium on Hearing** has been sponsored by the *Rijksuniversiteit Groningen*, the *Royal Netherlands Academy of Arts and Sciences*, the *Heinsius-Houbolt Fund*, the Foundation *Steun het Gehoorgestoorde Kind*, the Foundation *Keel-, Neus- en Oorheelkunde*, the Foundation *Groninger Universiteitsfonds*, and the companies *Tracor Europa* and *Bernafon Nederland B.V.*

Colophon:
Text layout by Alja Mensink.
All authors used word processors to develop their papers, and submitted the manuscripts either on diskette or through international computer net using the EARN-link in Groningen. The text files, produced with a dozen different word processors, were translated using R-DOC/X™ (Advanced Computer Innovations) or, in a few cases, 'manually', to PC-Write™ (Quicksoft). PC-Write version 2.71 was tailored to optimize printout on a Hewlett-Packard LaserJet™ II printer, using HP's K-(Times Roman™ and mathematical) and U-(Helvetica™)-fontcartridges. hd.

Previous symposia:

1. 1969: Driebergen, Netherlands. *Frequency Analysis and Periodicity Detection in Hearing*. Edited by R.Plomp and G.F.Smoorenburg (Sijthoff, Leiden) 1970.
2. 1972: Eindhoven, Netherlands. *Hearing Theory*. Edited by B.L.Cardozo (I.P.O., Eindhoven).
3. 1974: Tutzing, F.R.Germany. *Facts and Models in Hearing*. Edited by E.Zwicker and E.Terhardt (Springer Verlag, Berlin).
4. 1977: Keele, Great Britain. *Psychophysics and Physiology of Hearing*. Edited by E.F.Evans and J.P.Wilson (Academic Press, London).
5. 1980: Noordwijkerhout, Netherlands. *Psychophysical, Physiological and Behavioural Studies in Hearing*. Edited by G.van den Brink and F.A.Bilsen (Delft University Press, Delft).
6. 1983: Bad Nauheim, F.R.Germany. *Hearing- Physiological Bases and Psychophysics*. Edited by R.Klinke and R.Hartmann (Springer Verlag, Berlin).
7. 1986: Cambridge, Great Britain. *Auditory Frequency Selectivity*. Edited by B.C.J.Moore and R.D.Patterson (Plenum Press, New York).

Basic Issues in Hearing

Proceedings of the 8th International Symposium on Hearing

Paterswolde, Netherlands, April 5–9, 1988

Edited by

H. Duifhuis
Biophysics Department
University of Groningen
The Netherlands

J. W. Horst and H. P. Wit
Institute of Audiology
University Hospital Groningen
The Netherlands

1988

Academic Press
Harcourt Brace Jovanovich, Publishers
London San Diego New York Berkeley Boston
Sydney Tokyo Toronto