Perception & Psychophysics
1986, 40 (3), 183-187

Grouping of vowel harmonics by frequency modulation: Absence of effects on phonemic categorization

R. B. GARDNER and C. J. DARWIN
University of Sussex, Brighton, England

A mistuned harmonic makes a reduced contribution to the phonetic quality of a vowel. The two experiments reported here investigated whether the rate of frequency change over time of a harmonic influences whether it contributes perceptually to a vowel's quality. In these experiments, frequency-modulating one harmonic at a different rate or with a different phase from that used to modulate remaining harmonics of a vowel had no effect on the vowel's perceived category. These results are consistent with those of previous experiments and with the hypothesis that coherence of frequency modulation is not used to group together simultaneous frequency components into speech categories.

The harmonics of a periodic sound such as a voiced vowel have frequencies that are integer multiples of the fundamental. This fact can be exploited by perceptual mechanisms that group together sounds from different sources. For example, frequency components that deviate by more than around 3% from an integer multiple of the fundamental make a reduced contribution to the pitch of a complex tone (Moore, Peters, & Glasberg, 1985) or to the phonetic quality of a vowel (Darwin & Gardner, 1986). A difference of 8% completely excludes a harmonic from the calculation of pitch. In addition, formants that have a common harmonic spacing are more likely to be grouped together in determining phonetic quality than are those that have different harmonic spacings (Darwin, 1981; Scheffers, 1983).

It is less clear whether *dynamic* properties of pitch movement exert an independent effect on perceptual grouping, over and above the static effects just described. Will a component that is acceptably in tune be excluded because the trajectory of its frequency movement is different from that of its contemporaries?

Two simple types of pitch movement that are found in sung vowels, for example, are *vibrato* and *jitter*. In vibrato the frequency of the fundamental is modulated at a constant rate, whereas in jitter it fluctuates randomly. In both types of movement, harmonic relations between the frequency components are maintained—their modulation is coherent.

It is possible to synthesize signals in which the components have incoherent modulation (e.g., McAdams, 1984). One harmonic can be frequency-modulated at, say, a different rate from the others. Provided the depth of modulation is less than around 3%, the instantaneous frequencies of the harmonics would still be sufficiently close to integer ratios for static grouping mechanisms to treat all the harmonics as if they came from a common source, but there would be dynamic information available that could indicate that one of the harmonics did not belong with the others. In these experiments, we investigated whether such coherence of modulation can be exploited by the auditory grouping mechanisms responsible for determining the phonetic quality of a vowel.

The influence of incoherent vibrato and jitter on source assignment was recently studied by McAdams (1984). He found that complex tones consisting of 16 equal-amplitude components were perceived as being composed of multiple sources when one partial was modulated with incoherent jitter, so that its frequency modulation was inconsistent with that of the remaining harmonics. However, McAdams failed to find any effect of the coherence of modulation on the perceptual prominence of a target vowel whose fundamental was modulated at a different frequency from that of two other simultaneous vowels, or which was modulated against unmodulated background vowels.

McAdams's (1984) experiments indicate that coherence of frequency modulation of harmonics influences the number of sound sources but does not influence perceptual prominence. A similar dissociation, this time between the number of sound sources and their phonetic quality, has been found when different formants of a vowel or syllable have different pitches (Cutting, 1976; Darwin, 1981). Listeners can hear the phonetic quality given by their grouping together formants on different pitches while at the same time reporting that there is more than one sound source present.

A rather different approach was used by Bregman and Doehring (1984) to investigate whether rate of linear fre-

This research was supported by Grants GR/C 60099, GR/D 65930, and GR/C 8522 from the UK Science and Engineering Research Council. Early discussions with Al Bregman during his visit to the laboratory helped to crystallize the experiments. The authors' address is: Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England.

quency change over time (frequency slope) is used to group together simultaneously occurring frequency sweeps. To test how strongly a particular component was bound to others, they tested how easily it could form a different perceptual group with other components. They found that the middle one of three simultaneous frequency sweeps was more easily captured by a different perceptual stream when its frequency slope was not the same as that of the other simultaneously present sounds. If the slope of the sweep was parallel to the other tones, it was more easily captured when it did not form a simple harmonic relationship with them. This last result is consistent with the finding of Darwin and Gardner (1986) that mistuning a harmonic reduces its contribution to the phonetic quality of a vowel. However, the effect of the frequency modulation (FM) slope could also be based on mistuning; a different slope on the central component means that simple harmonic relationships are not maintained over time, introducing static mistuning. The time over which components with different slopes are sufficiently in tune (say ±8%) to be grouped together by purely static mechanisms will vary with the difference in slope. Bregman and Doehring's results are compatible with the hypothesis that simultaneous components are not grouped by virtue of a common slope of frequency change over time.

In summary, it is clear that a mistuned component is less well integrated into a complex than one that is in tune. The lack of integration results in subjects' hearing both multiple sound sources and a changed phonetic quality. But the only evidence that dynamic properties of frequency movement contribute to perceptual integration comes from McAdams's (1984) experiments on vibrato and jitter of harmonics, in which subjects judged the number of sources that they heard. There has been no evidence that the coherence of vibrato and jitter contributes to grouping as manifested by the perceived phonetic quality. In the experiments reported here, we looked for such evidence.

For vowels such as [I] and [e], which can be distinguished by the frequency of their low first formant (F1), the listener's computation of F1 is based on the relative amplitudes of individual resolved components of the vowel spectrum (Darwin, 1984b; Darwin & Gardner, 1985). The computation of F1 follows grouping processes which assign these components to the same source (Darwin, 1984a, 1984b). If common frequency modulation is used to group harmonics together, then a harmonic whose FM characteristics are inconsistent with the remaining harmonics of the vowel might be expected to be assigned to a different sound source and to contribute less to the assessment of vowel quality. The perceptual integration of the target harmonic into a vowel can be measured by phoneme boundary shifts produced by changes in perceived vowel color.

All of the present experiments involved manipulation of a harmonic close to the first formant in a series of vowels differing in first-formant frequency between [I] and [e]. Perceptual exclusion of a harmonic close to a formant peak gives a shift in the perceived first-formant frequency that can be detected in a categorization experiment as a change in the position of the phoneme boundary. If the frequency modulation of a harmonic causes it to be grouped out from the vowel, the [I]-[e] boundary should shift.

EXPERIMENT 1

The aim of this experiment was to determine whether incoherent modulation of a single harmonic of a vowel reduces the contribution of that harmonic to the vowel's phonetic quality. In order to do this, we used an [I]-[e] continuum, differing in F1, and we tested whether the phoneme boundary shifted when a harmonic near to F1 was modulated incoherently from the rest. As a control for simple mistuning effects, we included conditions in which the same harmonic was mistuned by a constant amount equal to its maximum mistuning under incoherent modulation. To calibrate the size of any grouping effect, we also included a condition in which the same harmonic was physically removed from the vowel.

If coherence of modulation is used to group harmonics for phonetic categorization, then we should find a phoneme boundary shift in the conditions in which the harmonic close to F1 is modulated incoherently relative to the remaining harmonics. This shift should be greater than that obtained with simple static mistuning. If the harmonic is being completely grouped out, the phoneme boundary shift should be as large as that found when the harmonic is physically removed from the vowel.

Method

Stimuli. Steady-state vowels were synthesized using additive sine-wave synthesis based on Klatt's (1980) cascade synthesizer. Klatt's published program was modified to produce the transfer function (after the initial spectrally flat pulse-train input) appropriate for a particular vowel. This transfer function was then evaluated at harmonic frequencies of a 125-Hz fundamental, and sine-waves of the appropriate amplitude and phase were added together to give the complete vowel. For harmonics of frequency-modulated vowels, the transfer function was evaluated at the instantaneous frequency for each sample point and the appropriate phase and amplitude values were derived. Vowel continua consisting of nine sounds of 500 msec duration with 16 msec rise/fall times were synthesized on a fundamental of 125 Hz. The sounds varied in the value of F1 from 375 to 543 Hz in equal 21-Hz increments, giving /I/-like sounds at low F1 values and /e/-like sounds at high F1 values. The values of the second, third, fourth, and fifth formants were 2300, 2900, 3800, and 4800 Hz, respectively. The bandwidths of the first three formants were 90, 110, and 170 Hz, respectively. The bandwidths of the fourth and fifth formants were set at 1000 Hz.

One experimental condition, the basic continuum, used no frequency modulation. For the other vowel continua, some or all of the harmonics were sinusoidally frequency modulated. The depth of modulation was always 2% (that is, 34 cents—within the range used by McAdams and well above detection threshold values) and the modulating waveform started in sine phase.

In one of the coherent modulation conditions all harmonics were modulated at a frequency of 6 Hz; in the other, all harmonics were modulated at a frequency of 10 Hz.

In the incoherent modulation conditions, the 500-Hz harmonic was chosen to receive different modulation frequencies from the other harmonics, because our previous experiments showed clear effects on the phoneme boundary of the perceptual removal of this harmonic from the vowel (e.g., Darwin, 1984a, 1984b). In two conditions, the 500-Hz component was frequency modulated at 6 and 10 Hz while the other harmonics were modulated at 10 and 6 Hz, respectively. In another condition, the 500-Hz component was modulated at 10 Hz against an unmodulated background; in another, it was unmodulated against a 6-Hz background.

Two mistuned conditions were also included, in which the unmodulated 500-Hz component was mistuned by ±10 Hz against an unmodulated background. These values corresponded to the greatest frequency deviation of the modulated 500-Hz condition and acted as a control for instantaneous mistuning in the conditions with a modulated target against an unmodulated background.

In addition, there was a continuum of unmodulated vowels with the 500-Hz component removed completely. This acted as a comparison for the size of the grouping effects. If an incoherently modulated component was grouped out completely, its phoneme boundary should approach that for this condition.

Procedure. To determine the phoneme boundary in a particular condition, only seven members of the continuum were used, to reduce the size of the experiment. The particular range chosen for each continuum was based on previous experiments. Each continuum member was repeated 10 times, giving a total of 70 stimuli per condition across the 10 conditions—a grand total of 700 trials, presented in quasi-random order. The range of a particular continuum (cf. Brady & Darwin, 1978) would not influence the results, because all the conditions were randomized together. Twelve subjects with normal hearing were used, each carrying out the complete experiment in one session. Stimuli were presented in random order. The subjects were seated in a sound-proofed cubicle. They listened to the sounds diotically over Sennheiser 414 headphones at 72 dB SPL on-line from a VAX-11/780 computer via an LPA-11K at a sampling frequency of 10 kHz; the sounds were low-pass filtered at 4.5 kHz and 48 dB/octave. Subjects responded on a VDU keyboard, pressing the *i* key for /I/ sounds and the *e* key for /E/ sounds. The subjects were free to repeat each sound as often as they liked. Trials followed keypresses after 1 sec. Trial numbers were displayed and subjects could take a rest at any time. The subjects received a practice session before the experimental session. The entire session lasted about half an hour.

The computer scored the data on-line and fitted a probit function to each individual subject's data for each continuum. The individual phoneme boundaries were taken as the 50% point of the probit function, expressed in terms of the F1 value used to program the synthesizer.

Results

Mistuned conditions and removed condition. The mistuned conditions acted as a control against which any phoneme boundary shift found for the modulated conditions must be compared. On the basis of previous work (Darwin & Gardner, 1986), we would expect mistuning by ±10 Hz, as used here, to give no significant shift in the boundary. That expectation was confirmed. The boundaries for the +10 Hz and −10 Hz mistuned conditions were both 442 Hz. These did not differ significantly from the original boundary value of 446 Hz (individual *t* test, $p > .05$).

Removing the 500-Hz component completely gave an estimate of the maximum shift we would expect if the 500-Hz component were being perceptually completely grouped out of the vowel percept. We found the predicted large upward shift in the boundary to 484 Hz. An individual *t* test showed this shift to be significant at $p < .01$.

The effects of modulating all components coherently. Modulating all components coherently provided a control for any change in vowel quality that modulation itself might introduce. The filled symbols of Figure 1 show the boundary values for the coherent modulation conditions, that is, conditions in which the target and the background modulation frequencies were equal. Individual *t* tests showed that modulation of the target at frequencies of 6 Hz and 10 Hz gave phoneme boundaries that were not significantly different ($p > .05$) from that for the unmodulated (0 Hz target and background) condition or from each other.

The effects of incoherent modulation. Figure 1 also shows the boundaries for the various incoherent modulation conditions, which are represented by open symbols. Individual *t* tests showed that modulating the 500-Hz target at 10 Hz against an unmodulated background produced no significant shift ($p > .05$) in the boundary compared to that for the unmodulated condition. Neither was there any effect of leaving the target unmodulated against a 6-Hz background.

Modulating the target at a frequency different from that of the background also produced no significant boundary shifts. Thus the phoneme boundary for a 6-Hz target against a 10-Hz background did not differ significantly from that found when all the harmonics were coherently modulated at 10 Hz. Conversely, the boundary for the 10-Hz target and 6-Hz background condition did not differ from the boundary for the coherent 6-Hz condition.

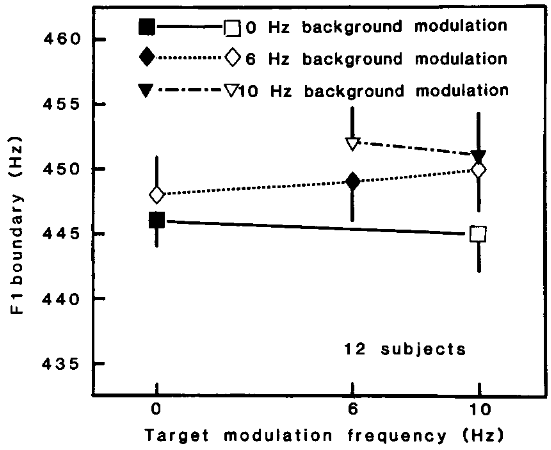

Figure 1. Experiment 1: Phoneme boundaries of vowel continua as a function of the modulation frequency of the 500-Hz target harmonic for different frequencies of background modulation. Filled symbols show the boundary values for coherent modulation. Open symbols show the boundaries for the various incoherent modulation conditions in which the target modulation frequency differed from that of the remaining background harmonics. Vertical bars are standard errors across 12 subjects.

In summary, no evidence was found for grouping by frequency modulation: Incoherently modulated targets did not produce the predicted upward shifts in the phoneme boundary.

Discussion

The results suggest that coherence of frequency modulation is not a necessary condition for the grouping together of harmonically related frequency components. Harmonic components whose FM characteristics did not match those of the background were fully integrated into the vowel spectrum. This was true not only for differences in modulation frequency between target and background, but also when the target was unmodulated against a modulated background and when it was modulated against an unmodulated background. One explanation of these results is that the differences in modulation frequency between the target and background harmonics were too small to be effective.

EXPERIMENT 2

This experiment was partially a replication of Experiment 1, but using a different range of target modulation frequencies against a 6-Hz background. In addition, a number of conditions were introduced in which the starting phase of the target modulation waveform was varied while its frequency was held constant at 6 Hz. This introduced a phase-based incoherence into the target modulation characteristics.

Method

Stimuli. The synthesis procedures and steady-state characteristics of the vowels were identical to those of Experiment 1. The original unmodulated vowel condition was again included in the experiment. The background modulation frequency was set at 6 Hz and conditions with target frequencies of 3, 6 (coherent), 12, 18, and 24 Hz were created. The depth of modulation was again 2% and the phase of the modulation waveform was 0°. Three further conditions were included, in which the target frequency was equal to the background frequency of 6 Hz but the starting phase of the target was varied. Values of 90°, 180°, and 270° were used.

Procedure. The experimental procedure was identical to that of Experiment 1 except that there were only nine conditions—a grand total of 630 trials. Eight subjects with normal hearing, 7 of whom had participated in Experiment 1, were used, each carrying out the experiment in one session.

Results

The effect of coherent modulation. The solid symbols in Figure 2 show the boundaries for the two coherent modulation conditions. As in Experiment 1, there was no significant difference between the boundary for the unmodulated condition (solid circle) and that for the condition in which all the harmonics were modulated coherently with a frequency of 6 Hz (solid triangle).

The effect of incoherent modulation. The open symbols in Figure 2 show the boundary values for the incoherent modulation conditions, in which the target frequency differed from that of the background. No differences were found between the boundaries for the coherent and incoherent modulation conditions.

Figure 3 shows that there was no effect of varying the phase of the target modulating waveform when its frequency was equal to that of the background (6 Hz).

Discussion

These results confirmed the findings of Experiment 1, that incoherent modulation of a harmonic had no effect on the integration of that harmonic into the vowel spectrum. This was true for differences between the modula-

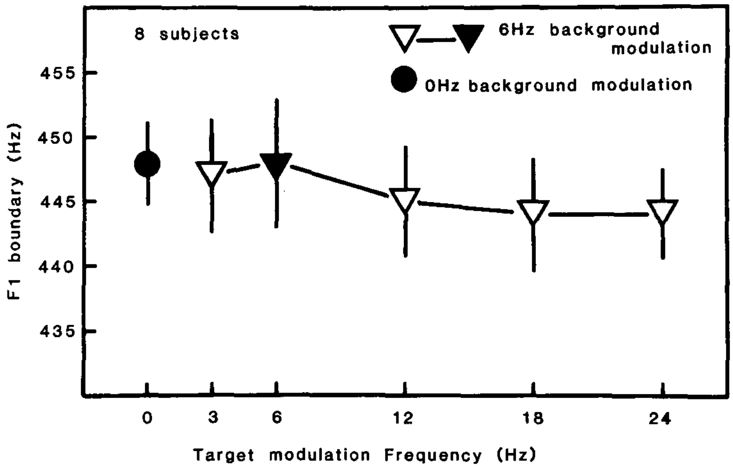

Figure 2. Experiment 2: Filled circle shows the boundary for the unmodulated vowel continuum (target and background modulation frequency equal to 0 Hz); filled triangle shows the boundary for vowels coherently modulated at a frequency of 6 Hz. Open symbols show the boundaries for the various target modulation frequencies of the incoherent modulation conditions, against a background modulation frequency of 6 Hz. Vertical bars are standard errors across 8 subjects.

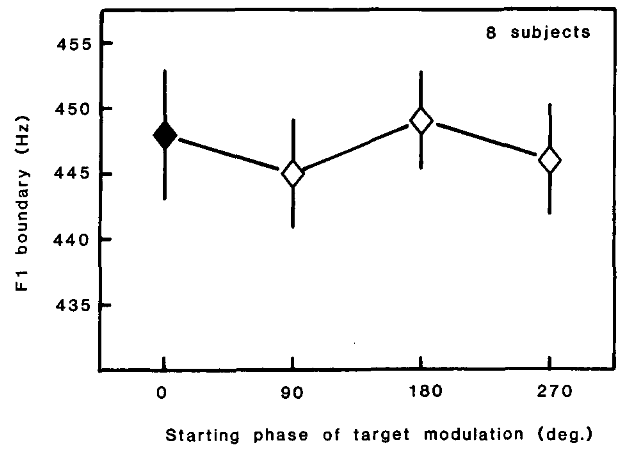

Figure 3. Experiment 2: Phoneme boundaries of vowel continua as a function of the starting phase of the modulation waveform of the 500-Hz target harmonic. Filled symbol shows the boundary value for coherent modulation. Open symbols show the boundaries for the various incoherent modulation conditions in which the starting phase of the target harmonic's modulating waveform differed from that of the background harmonics. Vertical bars are standard errors across 8 subjects.

tion frequency of the harmonic and that of the background of up to 2 octaves. There was also no effect of incoherence introduced by phase-shifting the modulating waveform of the target harmonic relative to that of the background harmonics, so that at a phase shift of 180° the frequency of the target rose while that of the others fell.

GENERAL DISCUSSION

The two experiments reported here have shown the following: (1) Modulation of a single harmonic of a vowel at a frequency different from that of the other harmonics does not affect the integration of that harmonic into the vowel spectrum (Experiments 1 and 2). (2) A change in the phase of the modulating waveform of a single harmonic of a vowel relative to that of the other harmonics does not influence the integration of that harmonic into the vowel spectrum (Experiment 2). (3) Modulation of a single harmonic against an unmodulated background of the remaining harmonics does not affect the integration of the harmonic into the vowel spectrum (Experiment 1). (4) No difference exists between the phoneme boundaries for unmodulated vowels and those for vowels in which all the harmonics are modulated coherently (Experiments 1 and 2).

In these experiments, we failed to find any evidence for the auditory system's use of coherence of modulation to group together resolved spectral components. This failure is unlikely to be due to insensitivity of the method employed; this method has proved extremely sensitive to the effects on phoneme boundaries of small amounts of added energy to a single harmonic (Darwin & Gardner, 1985), to the effects of mistuning a harmonic (Darwin & Gardner, 1986), and to the effects of making one harmonic start at a different time from another (Darwin, 1984a, 1984b). The failure is consistent with the proposal that incoherence in FM influences the number of sources heard but not the category perceived.

It is possible that other tasks may be able to reveal simultaneous grouping by frequency slope, since it is clear that we are able to detect the difference between coherent and incoherent modulation. Speech may not be the best paradigm to use in attempting to show such effects, since a substantial part of speech (all of voiceless speech and at least part of voiced speech) has excitation that is incoherent across different frequency regions. It might be more appropriate to use judgments of the timbre of a melodic instrument whose excitation is more consistently coherent than that of the voice.

REFERENCES

Brady, S. A., & Darwin, C. J. (1978). A range effect in the perception of voicing. *Journal of the Acoustical Society of America*, **63**, 1556-1558.

Bregman, A. S., & Doehring, P. (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. *Perception & Psychophysics*, **36**, 251-256.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, **83**, 114-140.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Quarterly Journal of Experimental Psychology*, **33A**, 185-207.

Darwin, C. J. (1984a). Auditory processing and speech perception. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 197-210). Hillsdale, NJ: Erlbaum.

Darwin, C. J. (1984b). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America*, **76**, 1636-1647.

Darwin, C. J., & Gardner, R. B. (1985). Which harmonics contribute to the estimation of the first formant? *Speech Communication*, **4**, 231-235.

Darwin, C. J., & Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America*, **79**, 838-844.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

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

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America*, **77**, 1853-1860.

Scheffers, M. T. (1983). *Sifting vowels: Auditory pitch analysis and sound segregation*. Doctoral thesis, Groningen University, The Netherlands.

(Manuscript received January 23, 1986;
revision accepted for publication June 17, 1986.)