Absence of effect of coherent frequency modulation on grouping a mistuned harmonic with a vowel

C. J. Darwin[a)] and Gregory J. Sandell
Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

(Received 5 October 1994; revised 13 January 1995; accepted 24 January 1995)

When a single harmonic close to the first formant frequency is mistuned by about 8%, that harmonic makes a reduced contribution to the vowel's first formant frequency as measured by a shift in the phoneme boundary along an $F1$ continuum between /ɪ/ and /ɛ/ [C. J. Darwin and R. B Gardner, J. Acoust. Soc. Am. **79**, 838–45 (1986)]. In the present experiments, phoneme boundaries along an /ɪ/–/ɛ/ continuum were measured for vowels differing in $F1$ whose fourth harmonic (500 Hz) was mistuned by 0, ±3, ±6, or ±9%. All the harmonics of a vowel (including the mistuned one) were given either no FM or coherent FM at a rate of 6 Hz and modulation depth of ±5%. The results replicated the previous findings, but found no evidence for coherent FM preventing the segregation of the mistuned harmonic from the vowel.

PACS numbers: 43.71.Es, 43.66.Jh, 43.66.Mk

INTRODUCTION

It is now well established that complex sounds can be grouped by a common harmonic structure for the purpose of calculating either their pitch (Moore *et al.*, 1985; Darwin and Ciocca, 1992) or their vowel quality (Darwin, 1981; Scheffers, 1983; Darwin and Gardner, 1986; Assmann and Summerfield, 1990). But, it is less clear what is the perceptual role of changes in fundamental frequency, or frequency modulation (FM), such as those provided by vibrato, or by a pitch contour in speech.

There are three possible perceptual roles that FM could play. First, it could directly help the identification of a complex sound's timbre or vowel quality by providing additional information about the spectral envelope of the complex sound. When the fundamental frequency is steady, information about the spectral envelope is only available at the harmonic frequencies, making it potentially difficult for the auditory system to estimate the position of low-frequency formant peaks in the spectral envelope where the individual harmonics are resolved. If, as in natural speech, the FM is such that frequency changes trace the spectral envelope, it might help the identification of a complex sound's timbre or vowel quality by providing additional information about the spectral envelope of the complex sound. However, there is surprisingly little evidence for this mechanism (Sundberg, 1977; McAdams and Rodet, 1988; Demany and Semal, 1990).

Second, the different patterns of movement that changes in fundamental frequency impose on the harmonics of periodic sounds could in principle indicate to the auditory system which harmonics originated from which source. There is now considerable evidence that sounds from two different sources cannot be segregated simply according to a different pattern of FM in each sound source. This somewhat surprising conclusion is true for different patterns of vibratolike FM failing to segregate one vowel from similar sounds (McAdams, 1989; Marin and McAdams, 1991; Summerfield and Culling, 1992) or to segregate a particular harmonic (Gardner and Darwin, 1986) or formant (Gardner *et al.*, 1989) from the perception of vowel quality. This failure may arise because listeners cannot detect the difference between coherent and incoherent FM across groups of harmonics occupying different frequency regions (Carlyon, 1991; Carlyon, 1994). There is some evidence that for larger fundamental frequency excursions, FM may play a small role in segregating voices (Chalikia and Bregman, 1993), but this possibility needs to be explored further.

Third, a common pattern of FM may help to group together those frequency components that share the common pattern. There is evidence that it does. When one voice in a chord is gradually given vibrato, it can be demonstrated to stand out as a coherent percept from the undifferentiated, unmodulated background (Chowning, 1980; McAdams, 1982). There is also evidence from recent experimental work on pitch perception. It exploits the fact that a slightly mistuned low-numbered, resolved harmonic of an otherwise harmonic complex sound will give a small pitch shift to the complex, whereas one that is mistuned by larger amounts will not contribute to the pitch of the complex (Moore *et al.*, 1985). This finding is consonant with the idea of a harmonic sieve which rejects from the calculation of pitch those resolved frequencies which are insufficiently close to a harmonic frequency (Duifhuis *et al.*, 1982). A mistuned harmonic of a complex will continue to contribute to the pitch of that complex at larger mistunings when both it and the rest of the complex have a common pattern of FM than when all the components are unmodulated (Darwin *et al.*, 1994). In other words, the tendency for mistuning to segregate a harmonic is reduced by that harmonic having a common FM with the rest of the complex.

On the other hand, there is also evidence that a common pattern of FM does not help to group together those frequency components that share the common pattern. Gardner *et al.* (1989) examined whether a single formant of a speechlike sound could be segregated from the other formants by a

[a)]E-mail: cjd@epvax.sussex.ac.uk

0001-4966/95/97(5)/3135/4/$6.00

difference in fundamental frequency and by differences in the fundamental frequency's FM. They used a four-formant stimulus in which all four formants gave the percept /ru/, whereas when the third formant was omitted the percept was /li/. They found that subjects heard the four-formant syllable change from /ru/ to /li/ as the fundamental frequency of the harmonics in the third formant region was made increasingly different from that in the other formants. This change was taken as evidence that the listeners were perceptually segregating the third formant from the other three. This tendency for the third formant of the composite syllable to segregate from the remaining four formants on the basis of a difference in fundamental frequency, was *not* reduced by a common pattern of FM on all the harmonics. In fact, there was a slight tendency in the opposite direction. Individual formants are not then grouped together on the basis of a common fundamental frequency FM.

The present experiment looks for a grouping effect of common FM using a different paradigm which has been used previously to demonstrate that a mistuned harmonic can be segregated from a vowel for the purpose of calculating vowel quality. It asks whether a common pattern of FM can help a mistuned harmonic to remain perceptually grouped with the other harmonics for the purpose of calculating vowel quality.

It is known that if a harmonic in the first formant region of a vowel is mistuned, it will make a reduced contribution to the calculation of the first formant frequency of the vowel (Darwin and Gardner, 1986). If common FM increases perceptual grouping, it should increase the contribution that the mistuned harmonic makes to the vowel percept.

The experiment is similar in design to the original study (Darwin and Gardner, 1986). The distinction between the steady-state, isolated vowels /ɪ/ and /ɛ/ can be made by changing only the first formant ($F1$) frequency. Changes in formant frequency change the relative amplitude of harmonics close to the formant peak. The phoneme boundary occurs for normal stimuli at an $F1$ frequency of about 450 Hz, so that an $F1$ below this value tends to yield an /ɪ/ percept and an $F1$ above this value tends to yield an /ɛ/ percept. When the vowels are synthesized on a fundamental frequency of 125 Hz, the third and fourth harmonics are similar in amplitude when the nominal (i.e., the synthesizer's) $F1$ frequency is around 450 Hz. If the fourth harmonic (500 Hz) is physically removed from the stimulus, the perceived $F1$ frequency is lower than the nominal frequency of 450 Hz, giving more /ɪ/ percepts. This lowering can be estimated in an identification experiment by measuring the phoneme boundary between /ɪ/ and /ɛ/ as a function of the nominal $F1$ frequency. When the fourth harmonic is removed, the phoneme boundary shifts to a higher nominal $F1$ frequency.

This upward shift in the phoneme boundary can then be used to estimate the contribution that a mistuned harmonic is making to the vowel color. If a mistuned harmonic is perceptually completely removed from the vowel, then we would expect to find a similar upward shift in the phoneme boundary than if it had been physically removed. Our previous results (Darwin and Gardner, 1986) showed that both progressive mistuning and physical removal of the harmonic gave increased $F1$ phoneme boundaries (corresponding to a lower perceived $F1$ frequency).

The question at issue in the present experiment is whether this increase in $F1$ frequency is reduced when a common pattern of FM is imposed on all the harmonics. If it is, then the experiment will have provided evidence that in vowel perception, as in pitch perception, a common pattern of FM can increase the perceptual coherence of a complex sound.

I. METHOD

A basic continuum of seven vowels differing in $F1$ were synthesized by harmonic addition using the transfer functions described in Klatt (1980). Each vowel had 30 harmonics of a 125-Hz fundamental and three formants. The frequency for the first formant was varied in 21-Hz steps from 396 to 522 Hz, while the second and third formants were fixed at 2100 and 2900 Hz. From this basic continuum two sets of eight continua were derived. In the first set (NoFM) the fourth harmonic was either mistuned by ±3%, 6%, or 9% of its harmonic frequency (500 Hz), not mistuned at all, or physically absent. In the second set (FM) all frequency components (including the possibly mistuned fourth harmonic) had a frequency modulation of 6 Hz, with a depth of ±5% (Darwin *et al.*, 1994). Modulated components traced out the appropriate spectral envelope and so had some amplitude modulation. For the perceptual consequences of such envelope tracing, see Marin and McAdams (1991) and McAdams and Rodet (1988).

The sounds had a duration of 500 ms, including 5-ms rise/fall raised cosine ramps and were presented diotically through Sennheiser HD 414 headphones at a level of 58 dB SPL. The whole set of 112 vowels was heard a total of ten times in a pseudo-randomized order in a single session.

The subjects' task was to identify whether they heard each vowel as /ɪ/ as in "bit" or /ɛ/ as in "bet" by pressing the "i" or "e" key on a keyboard. Subjects were only allowed to hear a vowel once before making their decision. After the subject pressed a key, the program played the next trial after a 500-ms pause.

Prior to the experiment, listeners were instructed in the task and played example stimuli. Vowels with the first formant at the extreme positions (396 and 522 Hz) were played, and listeners confirmed that they recognized them appropriately. Additionally, sequences of vowels traversing the seven formant values in order were played, including examples with and without mistuning and frequency modulation. Fourteen subjects in all participated, but two were rejected since in at least one of the 16 conditions their phoneme boundary could not be estimated since it fell outside the range of the stimuli used in the experiment.

II. RESULTS AND DISCUSSION

The average identification functions for the 12 subjects are shown in Figs. 1 and 2 and are orderly with similar slopes. Phoneme boundaries (50% /ɪ/ identification) for each subject and each condition were calculated from the individual identification functions by fitting them with a tanh

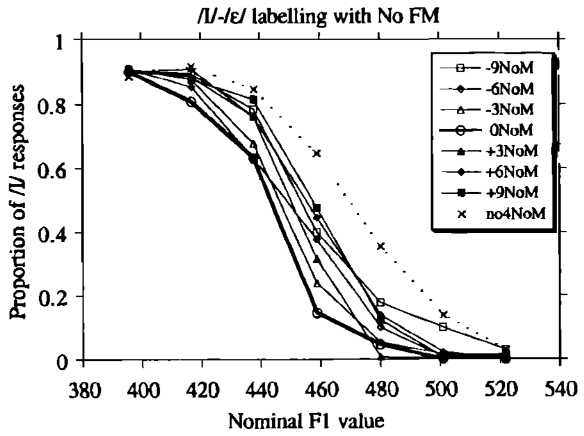

FIG. 1. Identification functions for an $F1$ continuum in which the fourth harmonic either has been mistuned by 0, ±3%, ±6%, or ±9% or has been physically removed (no4NoM condition).

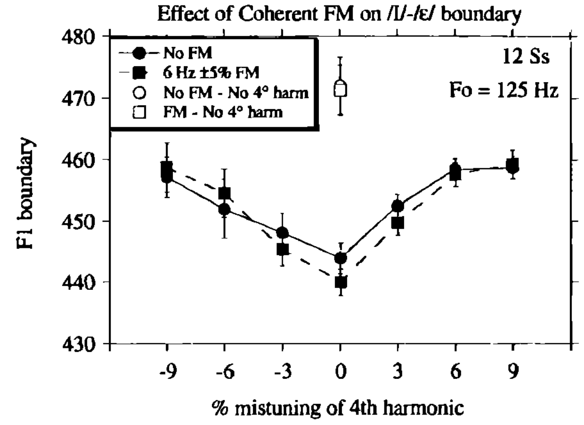

FIG. 3. Average phoneme boundaries across 12 subjects for $F1$ continua in which the fourth harmonic has been either mistuned by up to ±9% or has been physically removed. All components were either unmodulated (NoFM) or given coherent 6 Hz, ±5% FM. The error bars are ±1 standard error across the 12 subjects.

function; their mean values and standard errors across subjects are plotted in Fig. 3. The boundaries for the NoFM conditions replicate the findings of Darwin and Gardner (1986), showing that as the fourth harmonic is mistuned by up to 9%, the phoneme boundary moves to higher nominal $F1$ frequencies, approaching but not reaching the boundary value when the fourth harmonic is physically absent. These results are compatible with progressive mistuning, causing the fourth harmonic to be progressively removed from the calculation of vowel quality.

A repeated measures analysis of variance was carried out on the phoneme boundaries in each condition (except the No4 conditions) for each subject, with factors: presence/absence of FM (2) by levels of mistuning (7). It showed a highly significant overall effect of mistuning ($F_{6,66}=10.7$, $p<0.0001$) replicating the Darwin and Gardner result.

There was no main effect of FM/NoFM ($F_{1,11}=0.3$, $p>0.5$) nor any interaction of mistuning with FM/NoFM ($F_{6,66}=1.2$, $p>0.3$). This experiment thus provides no evidence that common FM helps to group a mistuned harmonic into a vowel for the purpose of calculating vowel quality. The small difference that we do find occurs in the opposite direction to that predicted from our previous experiments on the effect of FM on pitch perception (Darwin *et al.*, 1994). In those experiments, a mistuned component was more likely to be incorporated into the pitch percept under common FM than under no FM, particularly at large mistunings. If the same tendency were apparent in the present experiment, where the subjects' task is vowel identification rather than pitch perception, FM should serve to bind the mistuned component into the vowel percept, overcoming the tendency of the mistuned component to segregate from the vowel percept. Mistuning should therefore lead to less of a change in the phoneme boundary in the FM conditions than in the NoFM conditions. In Fig. 3 the FM curve should be shallower than the NoFM. But there is in fact an insignificant tendency in the *opposite* direction; the phoneme boundary function for the FM conditions is, if anything, steeper either side of zero mistuning than is that from the NoFM condition. A similar, and weakly significant, tendency was found in an earlier pilot experiment (Darwin and Sandell, 1994).

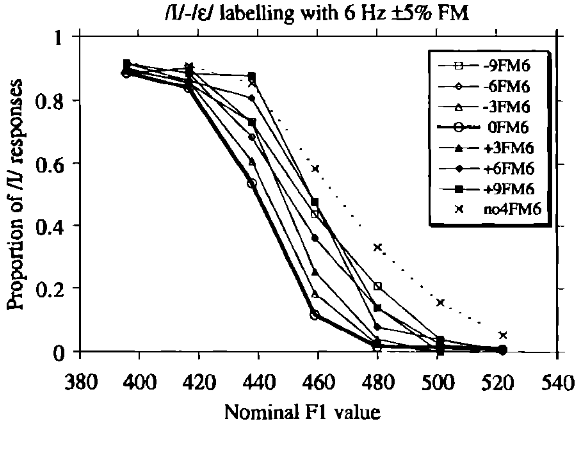

FIG. 2. Identification functions as in Fig. 1 but with all components being frequency modulated at a rate of 6 Hz and a depth of ±5%.

There is then no evidence in this experiment for common FM serving to group together the harmonics of a vowel for the purpose of calculating vowel quality. The only substantial difference between the stimuli in the present experiment and in the pitch experiment is that the pitch experiment used sounds that had a flat spectrum, whereas the present experiment, necessarily, used vowel-like stimuli. It is unlikely that this difference is responsible for the different use of FM. We have recently shown that such differences in spectral envelope are not responsible for differences in the use of onset-time as a grouping cue in vowel perception and pitch perception (Hukin and Darwin, 1995).

In summary, the present experiment has found no evidence that a common FM at 6 Hz and a width of ±5% can increase grouping for vowel perception. Vowel perception thus behaves differently from pitch perception, where the same depth and frequency of coherent FM *did* allow a mistuned harmonic to contribute more to the pitch of a complex tone than when there was no FM.

This result adds to the evidence that auditory grouping cues may vary in their effectiveness depending on the subject's perceptual task. We have recently presented evidence that onset asynchrony is more effective at segregating a harmonic from the calculation of vowel quality than it is at segregating the same harmonic from similar sounds for the calculation of pitch (Hukin and Darwin, 1995). Such findings have implications for the relationship between auditory grouping mechanisms and subsequent perceptual mechanisms involved in vowel or pitch perception. They argue against a simple model of grouping in which subsequent processes all share a common grouped output from primitive grouping mechanisms. Rather, they argue for different subsequent processes allocating different weights to the grouping resulting from the application of different primitive grouping cues. It is not yet clear whether the differences that we have found between pitch and vowel processing are specific to speech or whether they would also apply to the perception of nonspeech timbres such as musical instruments.

ACKNOWLEDGMENTS

This work was supported by BBSRC Grant GR/J66874 to the first author. John Culling kindly provided the software for the vowel synthesis and Paul Russell and Rob Hukin experimental software.

Assmann, P. F., and Summerfield, A. Q. (1990). "Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Carlyon, R. P. (1991). "Discriminating between coherent and incoherent frequency modulation of complex tones," J. Acoust. Soc. Am. **89**, 329–340.

Carlyon, R. P. (1994). "Further evidence against an across-frequency mechanism specific to the detection of frequency modulated (FM) incoherence between resolved frequency components," J. Acoust. Soc. Am. **95**, 949–961.

Chalikia, M. H., and Bregman, A. S. (1993). "The perceptual segregation of simultaneous vowels with harmonic, shifted or random components," Percept. Psychophys. **53**, 125–133.

Chowning, J. M. (1980). "Computer synthesis of the singing voice," in *Sound Generation in Wind, Strings, Computers*, edited by J. Sundberg (Royal Academy of Music, Stockholm).

Darwin, C. J. (1981). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. A**33**, 185–208.

Darwin, C. J., and Ciocca, V. (1992). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.

Darwin, C. J., Ciocca, V., and Sandell, G. R. (1994). "Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic," J. Acoust. Soc. Am. **95**, 2631–2636.

Darwin, C. J., and Gardner, R. B. (1986). "Mistuning a harmonic of a vowel: grouping and phase effects on vowel quality," J. Acoust. Soc. Am. **79**, 838–845.

Darwin, C. J., and Sandell, G. J. (1994). "Effect of coherent frequency modulation on grouping the harmonics of a vowel," J. Acoust. Soc. Am. **95**, 2964–2965.

Demany, L., and Semal, C. (1990). "The effect of vibrato on the recognition of masked vowels," Percept. Psychophys. **48**, 436–444.

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.

Gardner, R. B., and Darwin, C. J. (1986). "Grouping of vowel harmonics by frequency modulation: absence of effects on phonemic categorization," Percept. Psychophys. **40**, 183–187.

Gardner, R. B., Gaskill, S. A., and Darwin, C. J. (1989). "Perceptual grouping of formants with static and dynamic differences in fundamental frequency," J. Acoust. Soc. Am. **85**, 1329–1337.

Hukin, R. W., and Darwin, C. J. (1995). "Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification," to be published in Percept. Psychophys.

Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Marin, C. M. H., and McAdams, S. (1991). "Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width," J. Acoust. Soc. Am. **89**, 341–351.

McAdams, S. (1982). "Spectral fusion and the creation of auditory images," in *Music, Mind and Brain: The Neuropsychology of Music*, edited by M. Clynes (Plenum, New York), pp. 279–298.

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

McAdams, S., and Rodet, X. (1988). "The role of FM-induced AM in dynamic spectral profile analysis," in *Basic Issue in Hearing*, edited by H. Duifhuis, J. W. Jorst, and H. P. Wit (Academic, London), pp. 359–369.

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**, 1853–1860.

Scheffers, M. T. (1983). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Gröningen University, The Netherlands.

Summerfield, A. Q., and Culling, J. F. (1992). "Auditory segregation of competing voices: absence of effects of FM or AM coherence," in *Auditory Processing of Complex Sounds*, edited by R. P. Carlyon, C. J. Darwin, and I. J. Russell (Royal Society, London), pp. 63–71.

Sundberg, J. (1977). "Vibrato and vowel identification," Arch. Acoust. **2**, 257–266.