Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony

C. J. Darwin and R. W. Hukin
Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

(Received 20 May 1997; accepted for publication 10 October 1997)

The two experiments reported here examine how an inter-aural time difference (ITD) interacts with two other cues, mistuning and onset asynchrony, in reducing the contribution of a single frequency component to the perception of a vowel's identity. Previous experiments have shown that although ITD is generally rather ineffective at segregating a simultaneous harmonic frequency component from a vowel, it can produce some segregation when listeners have already been exposed to the isolated segregated component. A difference in ITD increases segregation overall in experiment 1 where the to-be-segregated component can also have a different onset time from the remainder of the vowel, and experiment 2 shows a similar result when the to-be-segregated component is mistuned. However, segregation by ITD is present just as strongly on trials when there is neither mistuning nor a difference in onset-time as on trials where these additional cues are present. Segregation on trials when there is neither mistuning nor a difference in onset-time is however larger in the present experiment which mixed all conditions together than in similar trials in an earlier experiment that had a blocked design [C. J. Darwin and R. W. Hukin, J. Acoust. Soc. Am. **102**, 2316–2324 (1997)]. The results show that segregation by ITD increases when other more potent cues are present in the experiment.
[S0001-4966(98)00302-6]

PACS numbers: 43.66.Mk, 43.66.Pn, 43.66.Qp, 43.71.Es [JWH]

INTRODUCTION

In the normal environment, a single cue that is useful for the perceptual segregation of a sound source rarely occurs in isolation. Frequency components originating from one sound source will generally differ from those originating from another in a variety of ways: they may have different onset times, be part of different harmonic series, and come from different directions. Experiments that have varied these cues individually have shown that the auditory system can exploit at least some of these cues to perceptually segregate frequency components into different putative sound sources.

For example, there is clear evidence from experiments on vowel perception that mistuning a single harmonic (Darwin and Gardner, 1986; Darwin and Sandell, 1994), giving it a different onset time from the rest of the vowel (Darwin, 1984), or playing it to the opposite ear from the rest of the vowel (Hukin and Darwin, 1995) will substantially remove it from the perception of vowel identity. These cues also affect other tasks, suggesting that the segregation that they provide is a robust phenomenon (Carlyon, 1994; Darwin and Carlyon, 1995; Hill and Bailey, 1997). Such perceptual grouping does not occur so straightforwardly as a result of differences in interaural time difference (ITD). A large difference in ITD (c. $\pm 600\ \mu$s) is either ineffective or very weak, when it is the only cue, at segregating either simultaneous, formantlike noise bands into vowellike pairs (Culling and Summerfield, 1995) or in segregating an individual harmonic from the perception of the vowel identity of a simultaneous steady-state vowel (Hukin and Darwin, 1995; Darwin and Hukin, 1997).

However, a difference in ITD *can* be effective at enhancing the segregation of a harmonic from a vowel under conditions where the listener has been made aware of an appropriate separate sound source. Using the same subjects as are used in the present experiments, Darwin and Hukin (1997, exp. 1) looked at the ability of a difference in ITD to segregate a harmonic from the percept of a vowel's identity under two different types of presentation (measured by a change in the phoneme boundary between /ɪ/ and /ɛ/ along a F_1 continuum). Under *blocked* presentation subjects heard a block of trials in which one of the harmonics of a vowel had either the same or a different ITD from the remaining harmonics. A difference in ITD gave no change in the phoneme boundary, indicating no segregation of the harmonic from the vowel. However, when these trials were *mixed* with others in which the vowel was preceded by a tone corresponding to the to-be-segregated harmonic, then a difference in ITD did produce segregation, even in trials that lacked the preceding tone sequence. Subsequent experiments (Darwin and Hukin, 1997) showed that this across-trial facilitation required the presence of a tone that corresponded both in identity and ITD to the to-be-segregated tone.

The present experiments ask whether the use of a difference in ITD can also be facilitated by other, more integral cues to segregation. In particular it asks whether the effectiveness of a difference in ITD in segregating a harmonic from a vowel is increased when it has a difference in onset time from the remainder of a vowel, or when it is mistuned relative to the other harmonics in the vowel.

Evidence that the processing of ITD is not independent of perceptual grouping cues such as onset asynchrony and mistuning comes from experiments on the localization of

complex tones. The experiments exploit the result that listeners use consistency of ITD across frequency to localize a band-limited noise (Trahiotis and Stern, 1989) or tonal complex (Hill and Darwin, 1996). A 500-Hz tone with an interaural phase difference that leads on the left ear by three-quarters of a cycle (1.5 ms) will be heard on the right in isolation because of phase ambiguity and a preference of the binaural system for short ITDs. However, if it is presented as part of a broader-band tonal complex whose other components share the same (1.5 ms) ITD, the whole complex will be heard on the left, apparently because the binaural system gives weight to the consistency of (an albeit long) ITD across frequency. This across-frequency integration of ITD can, however, be disturbed either by mistuning or by varying the onset time of the 500-Hz component. A few percent (c. 3%) mistuning, or a few tens of milliseconds (c. 40 ms) delay in onset time is sufficient to perceptually segregate the component from the complex, and to cause it to be heard back towards its original location on the left (Hill and Darwin, 1996). These results argue for the subjective location of a sound being determined after some perceptual grouping has occurred.

The present experiments use vowel identification rather than subjective location as a measure of segregation. Our previous experiments have indicated that grouping by ITD can be influenced by other grouping cues, such as onset asynchrony and mistuning, in two ways. These two ways make different predictions for experiments in which trials which either do or do not have the other grouping cues mixed together in the same block.

First, by analogy with Hill and Darwin's localization experiments, onset asynchrony or mistuning could influence grouping by ITD within a trial by providing a segregated tone which can then be localized separately by its different ITD. This mechanism predicts that ITD should be more effective at segregating a harmonic from a vowel on trials when the other grouping cues are present than on trials when the harmonic is both synchronous and exactly in tune. Informal listening to the sounds used in the present experiments indicated that it was very easy to hear a sufficiently mistuned or asynchronous harmonic that had a different ITD from the rest of the vowel as a separate sound in a distinct location.

Second, other grouping cues can influence grouping by ITD by the across-trial facilitation mechanism described earlier: the segregated and separately localized tone can facilitate segregation by ITD on trials when other grouping cues are not present. This across-trial mechanism predicts that ITD should also be effective on trials which do not have another grouping cue present.

The first experiment examines how segregation by ITD is influenced by onset asynchrony, and the second experiment examines how it is influenced by mistuning.

I. EXPERIMENT 1: ONSET ASYNCHRONY AND ITD

The first experiment examines how the ability of a difference in ITD to segregate a harmonic from the perception of a vowel's identity is influenced by that harmonic also having a different onset time from the rest of the vowel.

The experiment uses a well-established paradigm (Darwin, 1984) to measure the extent to which a 500-Hz component is segregated from a steady-state vowel. The segregation is measured as the shift in the phoneme boundary along a first formant (F_1) continuum between the vowels /ɪ/ and /ɛ/. Physical, and by inference perceptual, removal of the harmonic results in a phoneme-boundary shift to higher nominal F_1 values. In order to increase the size of phoneme boundary shift that removal of the 500-Hz component produces, conditions are also included in which the level of the 500-Hz component has been increased by 6 dB.

The 500-Hz component is given five different onset asynchronies ranging from 0 to 40 ms in order to provide some, but not complete, segregation by onset asynchrony. The 500-Hz component is also given either the same ITD as the rest of the vowel (which is always presented with an ITD leading on the left ear) or the opposite ITD.

A. Method

The basic stimuli were similar to those used in Hukin and Darwin (1995, exp. 2). On each trial subjects classified a single vowel as /ɪ/ or /ɛ/. The vowel was 56 ms in duration on a fundamental of 125 Hz and varied in F_1 frequency from 396 to 521 Hz in seven steps. Harmonic amplitudes were calculated from the source and transfer function of the Klatt (1980) synthesizer in serial mode with the first three formant bandwidths at 90, 110, and 170 Hz and the second and third formant frequencies at 2100 and 2900 Hz, respectively. The 500-Hz component of the vowel was presented either with the same ITD as the rest of the vowel ($+666\ \mu$s, with the left ear leading) or with the opposite ear leading. The 500-Hz component started 0, 10, 20, 30, or 40 ms before the rest of the vowel. Another condition was run in which the 500-Hz component was physically absent (no 500-Hz). Nine subjects with normal hearing took two different blocks of trials on separate days. The 500-Hz component was given an additional gain of 6 dB in the second block of trials. Each block had 770 trials: 10 replications of $7F_1$ values$\times$11 conditions (no 500-Hz$+$5 onset asynchronies $\times$2 gains). Other experimental details were as in Hukin and Darwin (1995, exp. 2).

B. Results

Phoneme boundaries were estimated from each subject's data in each condition. Mean boundaries are shown in Fig. 1. The results replicate previous results on segregation by onset asynchrony and in addition show a clear effect of segregation by ITD.

Physically removing the 500-Hz 0-dB component increases the phoneme boundary by about 35 Hz and removing the 500-Hz 6-dB component increases the phoneme boundary by about 45 Hz. Giving the 500 -Hz component an onset asynchrony of up to 40 ms increases the phoneme boundary by up to about 20 Hz in both the 0- and $+6$-dB conditions ($F_{4,32}=21.4$, $p<0.0001$), indicating that onset asynchrony is partly removing the component from the calculation of vowel quality. Giving the 500-Hz component a different ITD from the rest of the vowel further increases the phoneme

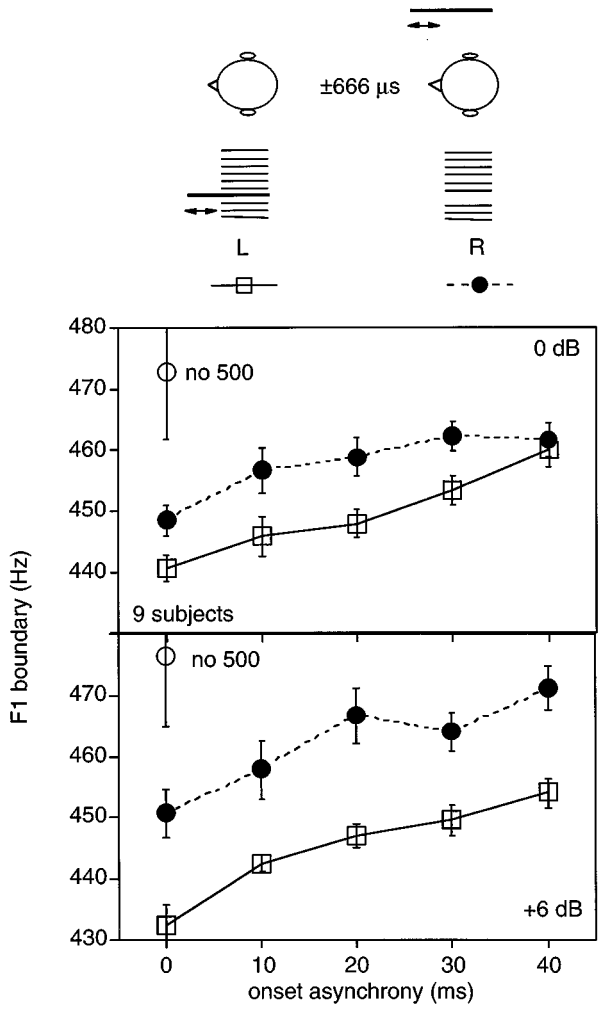

FIG. 1. Phoneme boundaries in experiment 1 along an /ɪ/–/ɛ/ continuum, for vowels with the 500-Hz component differing in onset asynchrony and in ITD from the rest of the vowel. Stimulus conditions are illustrated in the upper part of the figure, with frequency components placed on the ear that had the ITD lead. In the L condition all components had an ITD of $+666\ \mu$s, with the left ear leading. In the R condition the 500-Hz component was given an ITD of $-666\ \mu$s. In the lower panel the 500-Hz component has been given an additional gain of 6 dB across all conditions.

boundary ($F_{1,8}=33.3$, $p<0.0005$). Although both the 17-Hz increase with ITD for the +6-dB condition ($F_{1,8}=27.9$, $p<0.001$) and the 8-Hz increase for the 0-dB condition are significant ($F_{1,8}=16.8$, $p<0.005$), the increase is larger for the +6-dB condition ($F_{1,8}=8.4$, $p<0.02$). Overall, the size of the increase with ITD does not depend on onset asynchrony; in particular, it is not smaller at 0-ms onset asynchrony than at other asynchronies.

Across-trial facilitation of segregation by ITD was tested by comparing the size of the phoneme boundary shifts obtained here at 0-ms asynchrony with those found with the same subjects and the same stimuli presented in blocked conditions in Darwin and Hukin (1997, exp. 1). The shift is significantly larger in the present experiment than in the previous experiment, implying across-trial facilitation, for their +6-dB conditions ($F_{1,8}=8.1$, $p<0.025$) but not for their +0-dB conditions.

II. EXPERIMENT 2: MISTUNING AND ITD

The second experiment examines how the ability of a difference in ITD to segregate a harmonic from the perception of vowel's identity is influenced by that harmonic being mistuned relative to the rest of the vowel.

A. Stimuli

The stimuli for this experiment were similar to those of Experiment 1 except that the vowel duration was increased to 200 ms, all components were synchronous, and the 500-Hz component was mistuned by 0%, ±1%, ±2%, ±3%, and ±4%. The longer duration of 200 ms was chosen to be the same as that used in the previous vowel experiment on mistuning (Darwin and Gardner, 1986). The amplitude of the mistuned harmonic was held constant when it was mistuned. Again the 0- and +6-dB conditions were run in that order on separate days.

B. Results

Phoneme boundaries were estimated for each subject's data in each condition. Mean boundaries are shown in Fig. 2. The results replicate previous experiments on segregation by mistuning and in addition show a clear, though overall small, effect of segregation by ITD. Physically removing the 500-Hz component again increases the phoneme boundary by about 35 Hz for the 0-dB condition and about 45 Hz for the +6-dB condition. Mistuning the 500-Hz component generally increases the phoneme boundary (quadratic trend $F_{1,8}=37.0$, $p<0.002$), although the increase is more marked for the +6-dB condition than for the 0-dB ($F_{8,64}=4.4$, $p<0.01$) and for negative than for positive mistunings (linear trend $F_{1,8}=36.5$, $p<0.002$). The phoneme boundary is further increased when the 500-Hz component has a different ITD from the rest of the vowel ($F_{1,8}=67.0$, $p<0.0001$). This increase is significant for both the +6-dB condition (8 Hz, $F_{1,8}=45.2$, $p<0.001$) and for the 0-dB condition (4 Hz, $F_{1,8}=26.0$, $p<0.001$) and is larger ($F_{1,8}=7.3$, $p<0.05$) for the +6-dB condition than for the 0-dB condition. The size of the increase with ITD does not vary with mistuning; in particular, it is not smaller for zero mistuning.

Across-trial facilitation of segregation by ITD was tested by comparing the size of the phoneme boundary shifts obtained here at zero mistuning with those found with the same subjects and similar, though shorter, stimuli presented under blocked conditions in Darwin and Hukin (1997, exp. 1). The shift is significantly larger in the present experiment than in the previous experiment for their +6-dB conditions ($F_{1,8}=5.4$, $p<0.05$) but not for their +0-dB condition.

C. Discussion

Both experiments have shown clear evidence of segregation by ITD which adds to but does not interact with the segregation due to onset asynchrony or mistuning. Segregation by ITD is present just as strongly on those trials when there is no onset asynchrony or mistuning, as on those when these additional cues are present. This result contrasts with the weak or lack of segregation found when ITD is the only

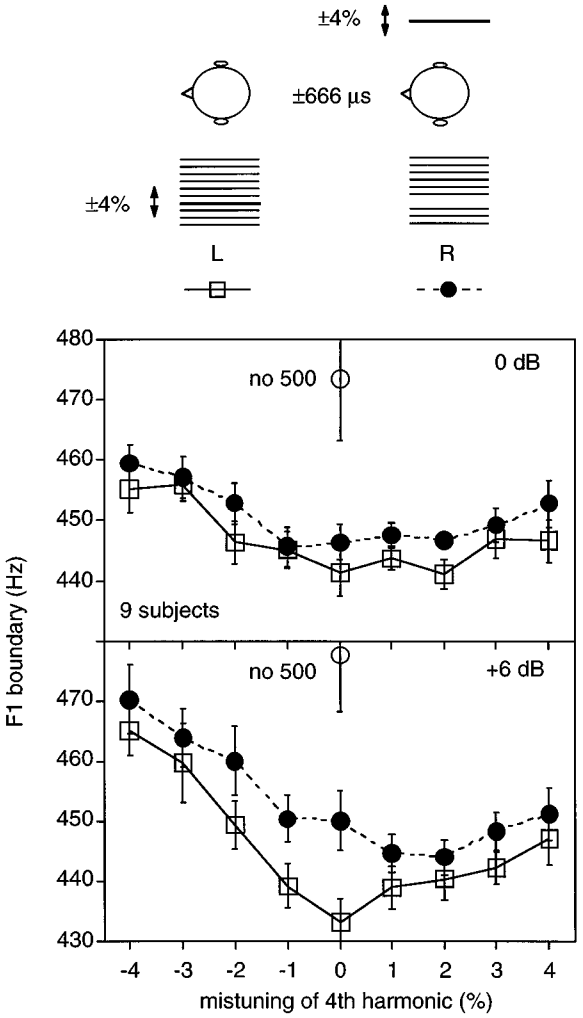

FIG. 2. Phoneme boundaries in experiment 2 along an /ɪ/–/ɛ/ continuum for vowels with the 500-Hz component mistuned and differing in ITD from the rest of the vowel. In the L condition all components had an ITD of $+666$ μs with the left ear leading. In the R condition the 500-Hz component was given an ITD of -666 μs. In the lower panel the 500-Hz component has been given an additional gain of 6 dB.

cue available to the listener in a block of trials (Culling and Summerfield, 1995; Hukin and Darwin, 1995; Darwin and Hukin, 1997).

It is likely that both the mechanisms proposed in the Introduction are occurring here. First, onset asynchrony and mistuning allow a harmonic from the vowel to be localized as a separate sound source. This may be a sufficient explanation for the additional removal by ITD of the harmonic from the vowel percept that we have found in trials where substantial amounts of onset asynchrony or mistuning are present. However, in order to explain why ITD also removes the harmonic from the vowel percept on trials when neither of these other cues are present, we must also appeal to the across-trial effect previously reported with mixed presentation by Hukin and Darwin (1995, exp. 2) and replicated in Darwin and Hukin (1997). In all these experiments listeners could hear, on some trials at least, an additional sound source corresponding to the tone that was to be segregated by ITD. In the previous experiments that additional sound source was explicitly present as a precursor tone; in the present experiments the tone could be heard out as a separate sound by virtue of its onset asynchrony or mistuning and localized on the opposite side to the rest of the vowel. The present results thus confirm that such across-trial effects can facilitate segregation by ITD, at least for the $+6$-dB condition (where phoneme boundary shifts are larger).

Segregation by ITD produces boundary shifts that are twice as large in combination with onset asynchrony than with mistuning. A possible reason for this is that listeners have a clearer image of the 500-Hz tone as a separate and distinctly localized sound source when it starts earlier than the rest of the vowel than when it is synchronous but mistuned. Although the average shifts produced by mistuning and onset asynchrony are comparable in these experiments, it is possible that the 500-Hz tone is more clearly localized to the opposite side when it leads the vowel. We have not formally investigated this possibility.

An entirely separate feature of the mistuning data is that negative mistunings produce more substantial shifts in the phoneme boundary (30 Hz) than do positive mistunings (particularly for the $+6$-dB condition). Fitting the data in Fig. 2 with a second-order polynomial decomposes the initially asymmetric function into a symmetric second-order component representing the U-shaped results expected from segregation due to mistuning, and a linear component representing the linear asymmetry. The linear component of the phoneme boundaries for the $+6$-dB L data has a change of about 20 Hz between $\pm 4\%$ mistuning. (The remaining second-order component for the $+6$-dB L data gives a shift of about 20 Hz for an absolute mistuning of 4%, which is comparable to that produced by an onset asynchrony of 40 ms.)

There are two possible reasons for the asymmetry. One is that it is a result of keeping the level of the mistuned 500-Hz constant rather than letting it follow the spectral envelope. Its constant level deviated from a 450-Hz formant envelope by about $+1$ and -3 dB for $+4\%$ and -4% mistunings, respectively. From the data of Fig. 2 a 6-dB increase in level shifts the boundary by about 10 Hz, so changes of this size could give a total asymmetry of about 7 Hz to the phoneme boundary, which is considerably smaller than that actually found. This explanation could be responsible for the slighter asymmetry seen in the 0-dB data. However, both of the previous experiments (Darwin and Gardner, 1986; Darwin and Sandell, 1994), that have shown phoneme boundary shifts with harmonic mistuning have varied level to maintain spectral envelope, but only the earlier study gave asymmetric data.

The second possible reason is that listener's estimates of the F_1 frequency are being directly influenced by the frequency of the mistuned harmonic. This explanation predicts a change of about 40 Hz, which is twice that observed, so neither explanation is clearly to be preferred.

III. SUMMARY

The two experiments reported here have shown that ITD can be used to perceptually segregate a harmonic from the calculation of a vowel's identity when it occurs in conjunction with other cues to perceptual segregation—onset asyn-

chrony and mistuning. Segregation by ITD also occurred in the present experiments on trials in which other segregation cues were absent. Since a difference in ITD has previously been shown to be ineffective at such perceptual segregation when it is the only cue in a block of trials, two mechanisms are proposed as an explanation for the present results.

First, onset asynchrony or mistuning can segregate a harmonic which may then in turn be localized on the opposite side to the vowel by virtue of a difference in ITD. Second, when listeners hear a separate sound source corresponding to the to-be-segregated harmonic, segregation by ITD is facilitated on other trials in the same experimental block.

In normal listening situations, where there are multiple cues to perceptual segregation, the first mechanism is likely to be the most important. However, the second mechanism does point to the possibility that ITD may be important in tracking a particular sound source over time.

ACKNOWLEDGMENTS

This research was supported by MRC Grant No. G9505738N. Bob Carlyon and Peter Assmann made helpful comments on the paper.

Carlyon, R. P. (**1994**). ''Detecting mistuning in the presence of asynchronous and asynchronous interfering sounds,'' J. Acoust. Soc. Am. **95**, 2622–2630.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J. (**1984**). ''Perceiving vowels in the presence of another sound: constraints on formant perception,'' J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *The handbook of perception and cognition, Volume 6, Hearing*, edited by B. C. J. Moore (Academic, London), pp. 387–424.

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 Hukin, R. W. (**1997**). ''Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity,'' J. Acoust. Soc. Am. **102**, 2316–2324.

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.

Hill, N. I., and Bailey, P. J. (**1997**). ''Profile analysis with an asynchronous target: Evidence for auditory grouping,'' J. Acoust. Soc. Am. **102**, 477–481.

Hill, N. I., and Darwin, C. J. (**1996**). ''Lateralisation of a perturbed harmonic: effects of onset asynchrony and mistuning,'' J. Acoust. Soc. Am. **100**, 2352–2364.

Hukin, R. W., and Darwin, C. J. (**1995**). ''Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel,'' J. Acoust. Soc. Am. **98**, 1380–1387.

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

Trahiotis, C., and Stern, R. M. (**1989**). ''Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase,'' J. Acoust. Soc. Am. **86**, 1285–1293.