Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity

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

(Received 17 January 1997; accepted for publication 20 June 1997)

The five experiments reported here examine the conditions under which sounds differing in their interaural time difference (ITD) are segregated for the purposes of perceiving a vowel's identity. Experiment 1 confirms previous findings that (i) a difference in ITD provides only a very weak cue for segregating a vowel's 500-Hz harmonic from the remainder of an isolated vowel; (ii) embedding the harmonic in a series of 500-Hz tones produces some segregation, which is enhanced if the harmonic and the tones differ in ITD from the rest of the vowel; and (iii) when these latter sounds are presented in the same block as isolated vowels, they facilitate segregation of the harmonic by ITD in the isolated vowels. The subsequent experiments show that this last effect, across-trial facilitation, is only produced by sounds which cue both the frequency and the ITD of the harmonic; either alone is insufficient. We also show that: (i) a single cue tone at the frequency of the harmonic is sufficient to facilitate the use of ITD in grouping; (ii) sequential organization by frequency proximity dominates over sequential organization by ITD when simultaneous sound sources are present; and (iii) the effectiveness of a cue tone can be abolished by capturing it into a synchronous harmonic complex. The experiments clarify the conditions under which ITDs contribute to the segregation of simultaneous sounds. © *1997 Acoustical Society of America.*
[S0001-4966(97)01610-X]

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

INTRODUCTION

This paper addresses the relationship between two different grouping cues: common frequency and interaural time difference (ITD). It is well-established that rapidly played tones that are similar in frequency will be more likely to form a single coherent stream or melody than will tones that are very different in frequency (Bregman and Campbell, 1971; van Noorden, 1975). Such coherence allows a pure tone that is part of a complex to be perceptually removed by a precursor tonal sequence with a consequent change in the complex's timbre (Bregman and Pinker, 1978), vowel quality (Darwin *et al.*, 1989), or pitch (Darwin *et al.*, 1995). Frequency proximity thus helps to define a sound source across time.

An alternative way to define a sound source across time is to exploit the fact that a sound source tends to come from a particular direction. Evidence for auditory grouping according to direction is more varied. Although, the identification of speech (Cherry and Taylor, 1954), or a tune (Deutsch, 1979) is severely disturbed by alternating the signal between the ears, the disturbance is reduced by adding another sound to the opposite ear (Schubert and Parker, 1956; Deutsch, 1979)—a finding which diminishes the usefulness of sequential grouping by common direction when multiple sound sources are present. A well-known example where grouping by frequency-proximity dominates over grouping by lateral position is in the opening bars of the last movement of Tschaikovsky's Sixth Symphony (Deutsch, 1982). Here the notes of the melody and of the accompanying ''second violin'' part alternate between the first and second violins (who would originally have sat on the conductor's left and right).

Recently, attention has been paid to the use in auditory grouping of specific cues to direction. In particular, Culling and Summerfield (1995) have demonstrated, somewhat counterintuitively, that interaural time differences (ITDs) cannot be used by the auditory system to segregate simultaneous sounds. They presented listeners with two simultaneous ''whispered'' vowels each of whose first two formants were represented by a pair of noise bands. They found that although listeners could use a common interaural level difference (ILD) to group together formants for vowel identification, they were unable to use a common ITD. Listeners were no better at identifying the vowel they heard on the left when the noise bands of the target vowel had a different ITD ($+390\ \mu$s) from those of the other vowel ($-390\ \mu$s) than when all four noise bands had the same ITD. The finding is the more surprising since low-frequency ITDs are the dominant cue for lateralizing complex sounds in azimuth (Wightman and Kistler, 1992).

The generality of the conclusion that listeners are unable to use ITDs to segregate simultaneous sounds was questioned by Hukin and Darwin (1995). They examined the effectiveness of an ITD difference in segregating the 500-Hz fourth harmonic of a vowel from the remaining harmonics. They measured the extent of the segregation through the shift in the phoneme boundary between the vowels /ɪ/ and /ɛ/ along a first-formant continuum (Darwin and Sutherland, 1984). This paradigm relies on the fact that the distinction between /ɪ/ and /ɛ/ can be cued by a change in first formant frequency ($F1$). Listeners can readily label sounds along a continuum differing in $F1$ as either /ɪ/ or /ɛ/; their phoneme boundary is then established at the 50% identification point. In our previous experiments this boundary normally occurs at an $F1$ of around 450 Hz. The perceived frequency of the first formant itself depends on the relative amplitude of nearby harmonics (Darwin and Gardner, 1985), so that if the amplitude of a harmonic close to the first-formant frequency is changed, the corresponding change in the perceived $F1$ can be detected by a shift in the /ɪ/-/ɛ/ phoneme boundary. The technique is sensitive to changes of around 1 dB in the level of harmonics close to $F1$. If the 500-Hz harmonic makes a reduced contribution to the vowel quality as a result of experimental manipulations, then the phoneme boundary should shift to higher nominal $F1$ values.

In partial support of Culling and Summerfield's conclusion that segregation of simultaneous sounds is not determined by ITD differences, Hukin and Darwin (1995) found that an ITD difference of $\pm 666\ \mu s$ between the 500-Hz component and the rest of the vowel gave a far smaller shift in the phoneme boundary than did an infinite ILD difference. The difference in ILD gave a shift in the phoneme boundary that was equivalent to a physical reduction in the level of the 500-Hz component of over 8 dB, whereas the ITD gave a reduction of about 2 dB. However, they did find that differences in ITD promoted segregation under slightly different stimulus conditions.

The conditions were derived from an earlier experiment (Darwin *et al.*, 1989) on grouping by frequency proximity which had shown that embedding a vowel in a short sequence of 500-Hz tones reduced the contribution of the 500-Hz harmonic to the vowel (measured as described above by a shift in the phoneme boundary). We will refer to this effect as within-trial segregation by tonal context. Hukin and Darwin (1995) showed that this segregation was greater when all the 500-Hz tones (surrounding tones and the harmonic of the vowel) were given a different ITD from the rest of the vowel ($\pm 666\ \mu s$). We will refer to this latter effect as within-trial segregation by ITD.

Hukin and Darwin also showed that when a block of trials contained trials that did contain the tone sequence mixed together with the original stimuli (that did not contain a tone sequence), then the latter sounds did show some segregation due to ITD (although recall that they showed no segregation when presented separately). We will refer to this effect as across-trial facilitation of segregation by ITD.

In the present paper we first replicate the experiment of Hukin and Darwin, using a group of subjects who will be used for all the experiments. We then show (experiments 2 and 3) that across-trial facilitation of segregation by ITD requires that the cue provides information about both the tonal quality and the lateral position of the harmonic that is to be segregated; neither is sufficient alone. That such across-trial facilitation can be produced by trials with only a single cue tone is shown in experiment 4. This experiment also demonstrates the important result that within-trial segregation by ITD depends only on the ITD of the cue tone, not on the ITD of the segregated harmonic. Experiment 5 shows that none of the above effects occur when the cue is itself grouped into a separate complex.

I. EXPERIMENT 1: TONE SEQUENCE AND ITD

The first experiment replicated the first experiment of Hukin and Darwin (1995). The replication was made for two reasons, first to ensure that the within-trial segregation by ITD and the across-trial facilitation of segregation by ITD that Hukin and Darwin reported were robust, and second to allow data from the same subjects to be compared across the remaining experiments.

A. Method

The method used in these experiments was the same as that used by Hukin and Darwin (1995). The contribution that the 500-Hz component of a vowel makes to its categorization as either /ɪ/ or /ɛ/ was estimated from the phoneme boundary along an $F1$ continuum. Physical or perceptual removal of the 500-Hz component from the vowel gives a shift in the phoneme boundary to higher nominal formant frequencies. In order to provide some calibration of the size of the effect, different continua had different physical gains of the 500-Hz component, in addition to those introduced by varying the $F1$ frequency.

The original vowel continuum contained six sounds whose first formant frequency varied between 379 and 500 Hz to give a percept that changed from /ɪ/ at low values of $F1$ to /ɛ/ at high values. The second- and third-formant frequencies were fixed at 2100 and 2900 Hz, respectively. The bandwidths of the three formants were kept constant at 90, 110, and 170 Hz. All sounds were synthesized on a fundamental frequency of 125 Hz and at a level such that the 500-Hz component of the member of the basic (0-dB gain) continuum with an $F1$ of 500 Hz was 60 dB SPL. The duration of the vowels was 56 ms (including 16-ms raised-cosine on- and off-ramps).

From this basic continuum were derived four sets of experimental continua in which the amplitude of the 500-Hz component was either 0 dB, +3 dB, +6 dB, or +9 dB relative to that in the basic continuum. Each condition in the following experiments thus consists of 24 different stimuli: 6 $F1$ values $\times 4$ gains of the 500-Hz component. In all conditions, the vowel components other than the 500 Hz were all presented binaurally with an ITD that led on the left ear ($\text{ITD} = +666\ \mu s$). The 500-Hz component was presented with either the same ITD as the other vowel components (condition L), or an ITD that led on the right ear (condition R: $\text{ITD} = -666\ \mu s$) to give a total of eight vowel continua. These two sets of four continua constituted the two uncued conditions.

Two further, cued conditions (LL, RR) were made by preceding each vowel with four, and following it with two, 500-Hz cue tones identical (including in level) to its 500-Hz component and with the same ITD of $\pm 666\ \mu s$. The cue tones were separated from each other and the vowel by 160 ms. The various conditions are illustrated in Fig. 1.

Subjects listened to the stimuli under two different modes of presentation. With blocked presentation the L and R conditions were given in one block of trials, and the LL and RR conditions in a separate block. With mixed presentation all four conditions were randomized together. With each mode of presentation subjects identified ten tokens of each stimulus in a pseudo-random order within blocks.

The sounds were synthesized by harmonic summation in real time with 16-bit resolution and at a sample rate of 44.1

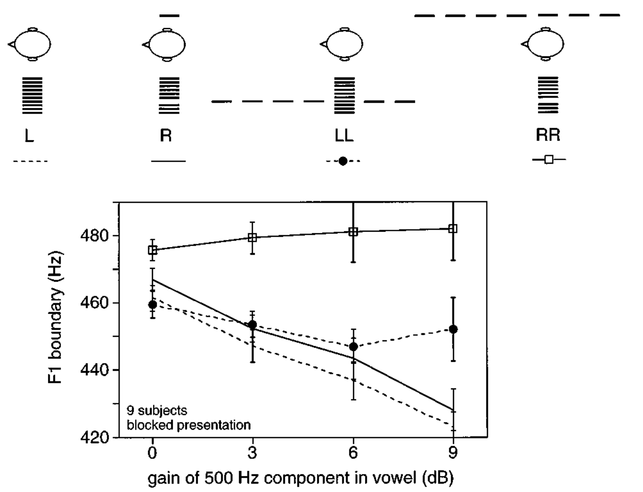

FIG. 1. Phoneme boundaries (with standard errors across subjects) along an /ɪ/-/ɛ/ continuum, with different additional gains on the 500-Hz component of the vowel for blocked presentation in experiment 1. With blocked presentation, the uncued conditions (L,R) are presented in one block, and the cued (LL,RR) in a separate block.

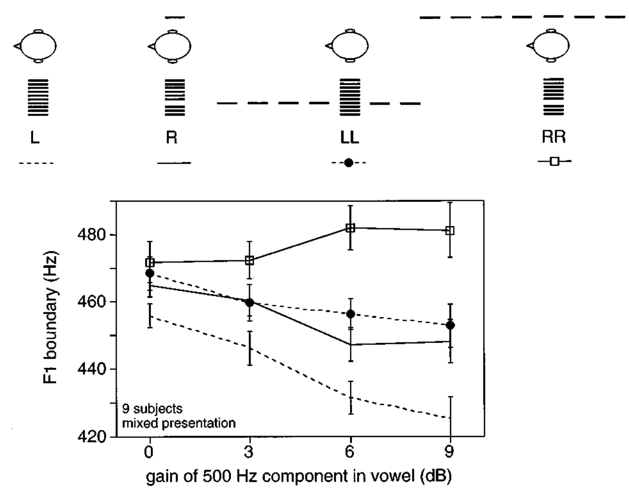

FIG. 2. Phoneme boundaries for mixed presentation in experiment 1.

kHz using custom software (Russell and Darwin, 1991) on a Digidesign Protools board attached to an Apple Macintosh IIfx, which also controlled the experiment. The vowel spectral envelopes were calculated using source and transfer functions taken from the Klatt synthesizer in serial mode (Klatt, 1980). The outputs of the Protools interface were connected to Tucker-Davis PA3 programmable attenuators which were used to set the overall level for the experiment. Subjects listened over Sennheiser HD414 headphones in a double-skinned IAC booth.

Subjects were told that they would hear a vowel in their left ear which could be either /ɪ/ as in ''pit,'' or /ɛ/ as in ''pet,'' and that they might also hear a tone in either ear, which they were to ignore. They signalled their response on each trial using the ''i'' and ''e'' keys on the Mac keyboard. Each trial followed 500 ms after the response to the previous one.

The same nine subjects took part in each experiment, all of whom had previous experience of similar vowel identification experiments. All were native speakers of British English with normal pure-tone thresholds over the range of frequencies of interest in this experiment. Two subjects had contributed data to experiment 2 of Hukin and Darwin (1995). Phoneme boundaries were estimated for each subject's data in each continuum using an automatically fitted rescaled tanh function.[1] All curve fits were checked by eye. Analysis of variance was carried out on these phoneme boundaries with SuperANOVA (Abacus Concepts, Inc.) using the Greenhouse-Geisser adjustment for correlation among repeated measures.

B. Results and discussion

Figures 1 and 2 show the mean phoneme boundaries and their standard errors across subjects. The results closely replicated the results of experiment 2 of Hukin and Darwin (1995). As expected, for both modes of presentation, phoneme boundaries in the L condition move to lower frequencies as the gain of the 500-Hz component increased, thus demonstrating the sensitivity of the paradigm to physical changes in the level of that component. An increase in the frequency of the phoneme boundary corresponds to a decrease in the effective level of the 500-Hz component.

1. Blocked presentation

Giving the 500-Hz component of the vowel a different ITD has no significant effect on the phoneme boundary (Condition L vs R: $F_{1,8}=3.8$, $p>0.05$). This result supports Culling and Summerfield's (1995) conclusion that a difference in ITD is an ineffective cue for segregation. As also found previously by Hukin and Darwin (1995), a tonal context produces two effects. First, it is effective at perceptually removing the 500-Hz harmonic from the vowel: the LL condition has significantly higher ($F_{1,8}=5.8$, $p<0.05$) and flatter—across changes in gain—($F_{3,24}=6.0$, $p<0.005$) phoneme boundaries than does the L condition (within-trial segregation by tonal context). Second, it allows ITD to produce additional segregation: The RR condition has much higher ($F_{1,8}=29.4$, $p<0.001$) phoneme boundaries than the LL condition (within-trial segregation by ITD).

2. Mixed presentation

The results for the L, LL, and RR condition are very similar with mixed and with blocked presentation. But, as found in the Hukin and Darwin's earlier experiment, the results for the R condition are different. Putting the L and R conditions in the same block of trials as the LL and RR conditions leads to the R condition having a significantly higher phoneme boundary than the L condition ($F_{1,8}=13.9$, $p<0.01$). This increase in the boundary in the R compared to the L condition is larger with mixed than with blocked presentation ($F_{1,8}=8.5$, $p<0.02$)—across-trial facilitation of segregation by ITD.

This experiment has replicated the three results reported in experiment 2 of Hukin and Darwin (1995). First, giving the 500-Hz component of a vowel a different ITD from the remaining components leads to no significant segregation of that component as measured by a shift in phoneme boundary. Second, embedding the vowel in a sequence of 500-Hz tones leads to some segregation of the 500-Hz harmonic (within-trial segregation by tonal context) which is substantially increased when all the 500-Hz tones are given a different ITD from the rest of the vowel (within-trial segregation by ITD). Third, mixing trials that contain the tone sequence with those that do not increases the segregation by ITD of the 500-Hz component from the vowel on trials without a tonal sequence: across-trial facilitation of segregation by ITD.

The following two experiments examine what aspect of the mixed presentation is responsible for inducing increased segregation by ITD in the R condition. Experiment 2 asks whether exposure to only the frequency of the 500-Hz tone is sufficient; experiment 3 asks whether exposure to the location of an additional, contralateral sound source is sufficient.

II. EXPERIMENT 2: FREQUENCY ONLY

This experiment asks whether the across-trial facilitation of segregation by ITD found by Hukin and Darwin (1995) and replicated in experiment 1 can be produced when subjects only have knowledge of the frequency of the to-be-segregated tone, rather than knowledge of both its frequency and its ITD.

These experiments found that a difference in ITD between the 500-Hz harmonic of a vowel and the remainder of the vowel was much more effective at removing that harmonic from the calculation of vowel quality in blocks of trials which also contained vowels embedded in a 500-Hz tone sequence that shared the harmonic's distinct ITD. The tone sequence on some trials thus shared both the frequency and the ITD of the to-be-segregated tone. Can a similar across-trial effect be obtained when the tone sequence only gives frequency, but not ITD information about the to-be-segregated tone? To answer this question experiment 2 repeats the mixed block part of experiment 1, but omits the condition in which the tone sequence has an ITD that leads on the right ear.

A. Stimuli

The experiment used mixed presentation of the L, R, and LL conditions of experiment 1. These three conditions were presented in a single mixed block of 720 trials (10 replications of 6 sounds from each continuum $\times 4$ levels of the 500-Hz component $\times 3$ conditions) to the same nine subjects of experiment 1.

B. Results and discussion

The mean phoneme boundaries and their standard errors are shown in Fig. 3. Replicating the previously found within-trial segregating effect of a tone sequence, phoneme boundaries are significantly higher ($F_{1,8}=11.1$, $p<0.02$) and flatter ($F_{3,24}=9.7$, $p<0.001$) in the presence of the tone sequence (condition LL) than without it (condition L). However, there is no evidence that the frequency information present in the tone sequence produces an across-trial facilitation of segregation by ITD. If there were, we would expect the R condition to have higher phoneme boundaries than the L, which it does not. Facilitation of segregation is significantly smaller here ($F_{1,8}=22.7$, $p<0.002$) than that found in the previous experiment when both frequency and ITD information was available from the tonal context. So, making subjects aware that there is a separate sound source present at 500-Hz during the same block of trials is not sufficient to induce segregation of the 500-Hz tone by ITD in the R condition. The conditions necessary for across-trial induction of segregation by ITD must therefore include some information about the location (or the ITD) of the additional sound source.

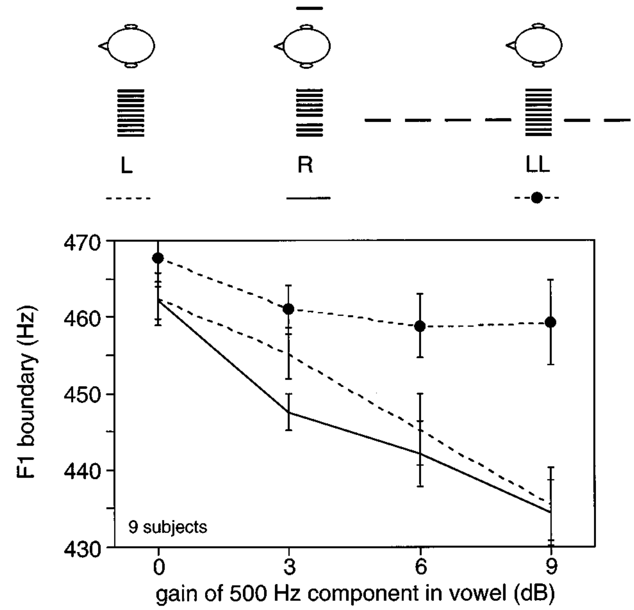

FIG. 3. Phoneme boundaries in experiment 2. The experiment used mixed presentation of a subset of conditions from experiment 1.

Is any sound with the same ITD a sufficient cue to give across-trial facilitation of segregation? The following experiment asks whether a narrow-band noise centered around the frequency of the tone can give such across-trial facilitation. Experiment 3 also asks whether such a narrow-band noise can produce the within-trial segregation either by context or by ITD found with tone sequences.

III. EXPERIMENT 3: NARROW-BAND NOISE

A. Method

The experiment was identical in design to experiment 1 except that the 500-Hz six-tone sequence used in the LL and RR conditions of experiment 1 was replaced by a similar sequence of six narrow-band noises. The noise was centered around 500 Hz with a bandwidth of 70 Hz. It consisted of harmonics 47–54 of a flat amplitude, random-phase signal with a fundamental of 10 Hz. The level of the noise in the 0-dB conditions was such that, when passed through a (fourth-order) gammatone filter centered on 500 Hz, it had the same energy as that of the 500-Hz component of a vowel in the original continuum with $F1=500$ Hz. Its level was incremented in steps of 3 dB with the level of the 500-Hz component of the vowel. The narrow-band noise thus pro-

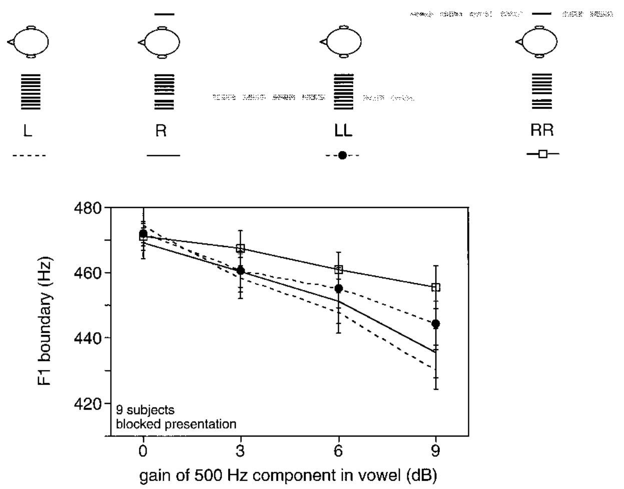

FIG. 4. Phoneme boundaries for blocked presentation in experiment 3. The experiment was similar to experiment 1 but used narrow-band noise as the cue.

vides similar energy in the auditory channel responding to a 500-Hz tone, but is different in timbre, and so is less likely to stream with the 500-Hz component of the vowel. The same nine subjects took the experiment.

B. Results

1. Blocked presentation

The phoneme boundaries for blocked presentation are shown in Fig. 4. As found in experiment 1, giving the 500-Hz component of the vowel a different ITD again has no significant effect on the phoneme boundary: The R condition's boundaries are not significantly higher ($F_{1,8}=0.3$, $p>0.5$) or flatter ($F_{3,24}=1.2$, $p>0.3$) than the L condition's. However, unlike experiment 1, a change in ITD also has no effect when the vowel is put in a sequence of narrow-band noise bursts: The RR condition is neither higher ($F_{1,8}=1.0$, $p>0.3$) nor flatter ($F_{3,24}=1.8$, $p>0.1$) than the LL. The sequence of noise bursts thus produces no within-trial segregation by ITD.

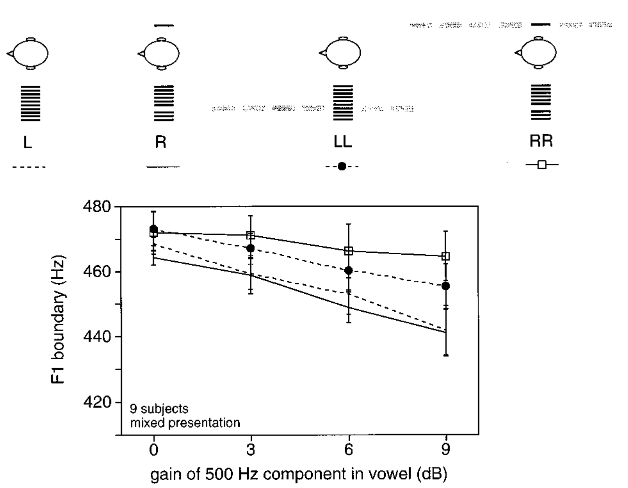

FIG. 5. Phoneme boundaries for mixed presentation in experiment 3. The experiment was similar to experiment 1 but used narrow-band noise as the cue.

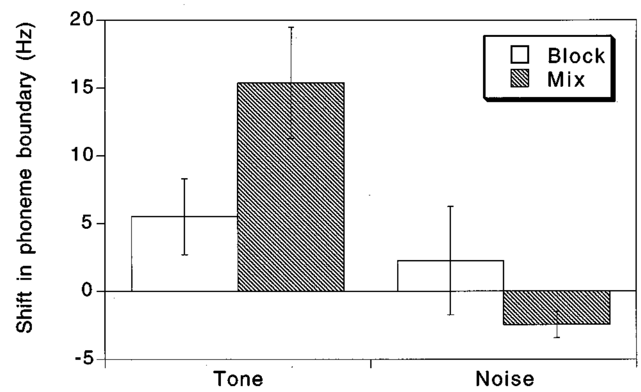

FIG. 6. Comparison of across-trial facilitation of segregation by ITD produced with mixed presentation by tone sequences in experiment 1 and narrow-band noise sequences in experiment 3. Each bar represents the difference in phoneme boundary between the R and L conditions averaged across the four levels of the 500-Hz component. The extent of across-trial facilitation is given by the difference between the mixed and the blocked conditions for each type of cue sequence.

2. Mixed presentation

The pattern of results with mixed presentation is very similar to that found with blocked presentation (Fig. 5). Although the difference in phoneme boundary between the L and R conditions is marginally significant ($F_{1,8}=6.3$, $p<0.05$), the size of the effect is small and it is in the opposite direction to that expected if segregation were occurring. A sequence of noise bursts leading on the opposite ear thus does not produce any across-trial facilitation of segregation by ITD. The tonal sequence in experiment 1 produces significantly more across-trial facilitation (averaging across level of the 500-Hz component) than does the noise sequence in experiment 3 ($F_{1,8}=18.04$, $p<0.005$), as illustrated in Figs. 6.

The noise sequence flattens the phoneme boundary function slightly: The LL condition gives flatter phoneme boundaries than the L ($F_{3,24}=4.8$, $p<0.02$). Overall, the noise context produces smaller changes in the phoneme boundaries than did the tonal context of experiment 1 ($F_{1,8}=8.7$, $p<0.02$). Since the small change produced by the noise context (although not that produced by the tonal context) is independent of ITD, it could be the result of peripheral adaptation to the 500-Hz noise band reducing the response to the 500-Hz harmonic of the vowel.

In summary, a sequence of 500-Hz narrow-band noise bursts is much less effective than a tone sequence at producing either within-trial or across-trial effects of ITD on grouping. Taking experiments 2 and 3 together, subjects need very specific information about the nature of an additional sound source before they show across-trial facilitation of segregation by ITD: Information about both the ITD and the specific tonal quality of the sound is necessary.

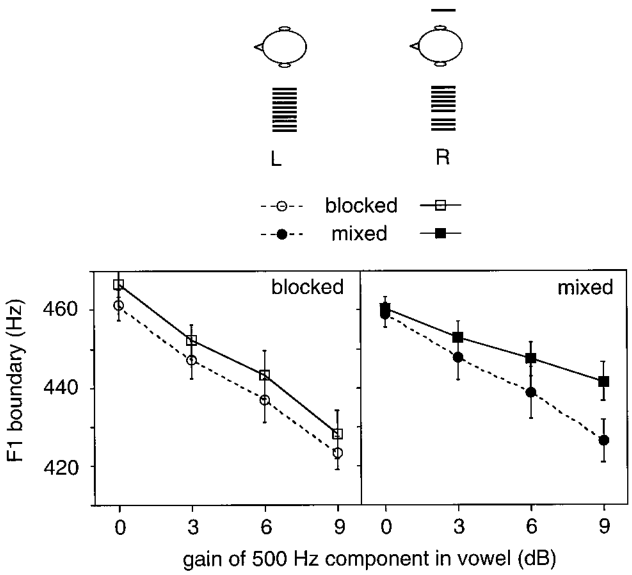

FIG. 7. Phoneme boundaries in the uncued conditions of experiment 4 with blocked and mixed presentation. The experiment was similar to experiment 1 but used a single tone as the cue.

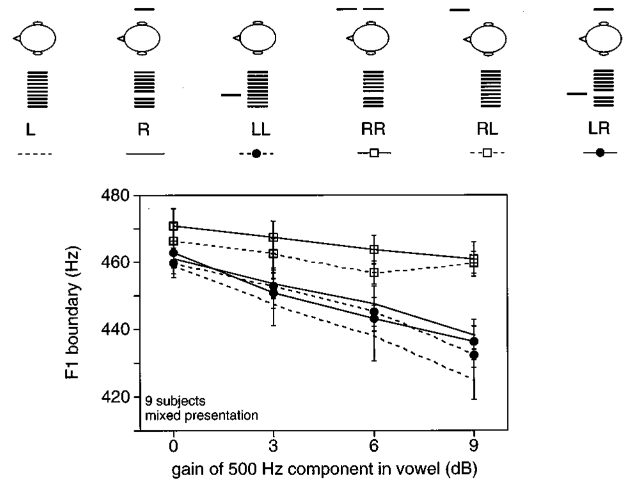

FIG. 8. Phoneme boundaries in the mixed conditions of experiment 4. The experiment was similar to experiment 1 but used a single tone as the cue.

IV. EXPERIMENT 4: SINGLE CUE

Experiment 4 differed from experiment 1 in two ways. First, the tonal context used was a single 500-Hz precursor tone rather than multiple tones. A single cue tone, although providing less sequential grouping than four tones, extends the generality of the effects that we have found. Second, conditions were included to investigate the interaction between grouping by frequency proximity and by ITD. In particular, the precursor 500-Hz tone was played with an ITD favoring one ear, while the 500-Hz harmonic of the vowel was played with an ITD favoring either the same or the opposite ear. Deutsch (1979) found that sequential grouping by frequency dominated over the ear of presentation for the identification of tunes whose notes alternated between the ears only when each note of the tune was accompanied by a drone note to the opposite ear. If a similar dominance of sequential grouping by frequency proximity over ITD occurs for our stimuli, then the extent to which the 500-Hz component is removed from the vowel should depend less on its location than on that of the precursor tone.

A. Method

The L and R conditions were identical to those of experiment 1. Four further, cued conditions (LL, LR, RL, RR) were made by preceding the vowel with a single 500-Hz precursor ''cue'' tone identical to its 500-Hz component but with an ITD of $\pm 666\ \mu$s that was independent of the ITD of the 500-Hz tone in the vowel. The first letter in the condition name refers to the leading ear for the cue tone, the second to the leading ear for the 500-Hz harmonic of the vowel. The cue tone ended 160 ms before the onset of the vowel. The various conditions are illustrated in Figs. 7 and 8.

The experiment was run in three blocks: one with ten repetitions of just the two uncued conditions (blocked L and R), and two mixed blocks, each with five repetitions of all six conditions (mixed L, R, LL, RL, LR, and RR). Each mixed block thus had 720 trials (5 replications $\times 6$ $F1$ values $\times 4$ levels levels of the 500-Hz component $\times 2$ lateralizations of the 500-Hz component $\times 3$ types of cue tone: no cue tone, $\pm 666\ \mu$s). The trials within each block were presented in different, pseudo-random orders.

B. Results and discussion

If the present cueing paradigm with a single precursor tone were as effective as experiment 1 was with four preceding and two following tones, we would expect to find: (i) within-trial segregation by tonal context (LL and RR conditions giving higher-frequency boundaries than L and R, respectively); (ii) within-trial segregation by ITD (RR condition giving higher-frequency boundaries than L); (iii) across-trial facilitation of segregation by ITD (R condition giving a greater increase in boundary frequency over the L condition with mixed than with blocked presentation).

In addition, the experiment asked whether sequential grouping by frequency can overrule grouping by ITD. The experiment included two conditions where the 500-Hz cue tone and the 500-Hz harmonic of the vowel had opposite ITDs (conditions LR and RL). If the cue tone can, by virtue of its proximity in frequency to the vowel harmonic, segregate the harmonic from the vowel regardless of their relative ITDs, then we might find (iv) similar, high phoneme boundaries in the RR and RL conditions on the one hand, and lower boundaries in the LL and LR conditions on the other.

All four effects were found in the present experiment, although, as found previously (Darwin *et al.*, 1989), the single cue tone produced weaker segregation than did a longer sequence.

1. Uncued conditions

Looking first at the results from the block of trials which consisted only of the uncued conditions (Fig. 7), we find, as expected from previous experiments, that giving the 500-Hz component of the vowel a different ITD has no effect on the phoneme boundaries: With blocked presentation condition R is no different from condition L ($F_{1,8}=3.8$, $p>0.05$).

Looking now at the results from the uncued conditions with mixed rather than blocked presentation, the R condition has significant higher boundaries than the *L* ($F_{1,8}=6.7$, $p<0.05$), a difference which increases with the gain of the 500-Hz component ($F_{3,24}=5.6$, $p<0.01$). Moreover, this relative increase is significantly bigger with mixed than with blocked presentation ($F_{3,24}=3.7$, $p<0.05$). These differences replicate with a single cue tone the across-trial facilitation of segregation by ITD from experiment 1 in which cues to the frequency and location of a separate sound source from other trials in the same experimental block allow the segregating effect of a difference in ITD to emerge.

2. Cued conditions

Figure 8 shows the results for mixed presentation. The within-trial segregation by tonal context produced by a single cue tone when all sounds have the same ITD is weak: the LL condition is not significantly different overall from the L condition ($F_{1,8}=2.0$, $p>0.01$). However, more specific *post hoc* contrasts (based on the expected pattern of results from earlier experiments) within this main effect show that the two higher gains of the LL condition have a higher boundary than those of the L condition (+6 dB: $p<0.02$; +9 dB: $p<0.002$). Overall weak segregation from a single preceding tone is compatible with the earlier finding (Darwin *et al.*, 1989) that the extent to which a harmonic is segregated from a vowel increases as the number of preceding tones increases upto four.

Preceding the vowel by a single cue tone having the same level and ITD as the 500-Hz component of the vowel, but a different ITD from the rest of the sounds, produces substantial changes in the frequencies of the phoneme boundaries. The phoneme boundaries are at a higher frequency for the RR condition than for the R ($F_{1,8}=18.2$, $p<0.005$). Moreover, the increase in boundary frequency is greater for the RR condition than for the LL condition ($F_{1,8}=11.1$, $p<0.02$). This last result extends to a single precursor tone the result of our previous paper (Hukin and Darwin, 1995) that differences in ITD produce within-trial segregation by ITD, enhancing the segregating effect of a tonal context on the harmonic of a vowel.

However, there is an additional, novel feature of the results of this experiment: The shift in the phoneme boundary in the cue tone conditions relative to the uncued conditions is determined more by the ITD of the cue tone ($F_{1,8}=16.5$, $p<0.005$), than by the ITD of the harmonic in the vowel ($F_{1,8}=5.6$, $p<0.05$). As a measure of the dominance of the cue tone's ITD over the harmonic's ITD, the RL condition has significantly higher ($F_{1,8}=10.8$, $p<0.02$) and flatter ($F_{3,24}=4.1$, $p<0.05$) phoneme boundaries than the LR condition.

The likely explanation for this last result is that sequential grouping by frequency proximity exerts a dominant influence in causing the two 500-Hz tones to group together (perhaps with the apparent location of the second being attributed to that of the first, cue tone). This dominance of the cue tone's location is likely to depend on the fact that the cue tone is presented alone (and so has a clear location), while the harmonic of the vowel is presented simultaneously with other harmonically related sounds and so has a more labile subjective location (cf. Hill and Darwin, 1996).

The irrelevance of the ITD of the 500-Hz harmonic to the extent to which it is segregated from the vowel by a cue tone reinforces Culling and Summerfield's conclusion that ITD is a very poor segregating cue for simultaneous sounds. Differences in ITD can however influence the segregating power (via frequency proximity) of isolated sounds presented at different times.

In summary, experiment 4 has shown:

(i) within-trial segregation by tonal context. A single cue tone at the frequency of the 500-Hz component of a vowel can produce some segregation of that component from a vowel when all sounds have the same ITD. This effect is only significant at higher levels of the 500-Hz component;

(ii) within-trial segregation by ITD. When the cue tone has a different ITD ($-666\ \mu$s) from the other sounds ($+666\ \mu$s), segregation of the 500-Hz harmonic from the vowel is substantial. An important new finding is that this segregation occurs regardless of the ITD of the 500-Hz harmonic;

(iii) across-trial facilitation of segregation by ITD. For the uncued conditions with mixed presentation (including trials with the single cue tone present in the same block of trials), there is significantly more segregation by ITD than with blocked presentation.

V. EXPERIMENT 5: COMPLEX SINGLE CUE

Experiment 5 acts as a control experiment for the results of experiment 4 being due to purely local effects around 500 Hz. The explanations that we have offered for the segregating effects of a tonal context have been based on principles of perceptual organisation. These principles are further invoked in experiment 2 to construct conditions where there should be no grouping by frequency proximity and so no tendency for the cue tone to segregate the 500-Hz harmonic from the vowel.

In this experiment we add four higher harmonics of 166.7 Hz synchronously to the 500-Hz cue tone. The 500-Hz cue tone should then form part of a single 166.7-Hz complex and not therefore be free to form a separate perceptual group with the 500-Hz harmonic of the vowel. A similar manipulation has been used to regroup the leading portion of a component of a complex to prevent it being removed, by virtue of its onset time, from the calculation of vowel quality (Darwin and Sutherland, 1984), pitch (Ciocca and Darwin, 1993), or profile analysis (Hill and Bailey, 1997). If the effects that we have found in the first experiment are due to mechanisms that are purely local to 500 Hz, then we would expect them also to appear in this experiment. If they are due to perceptual grouping, then they all should disappear.

A. Method

The experiment was similar to the mixed part of experiment 1, and used the same subjects. In addition to the cue tone, all conditions also contained a tonal complex that was simultaneous with the cue tone, but could be presented with

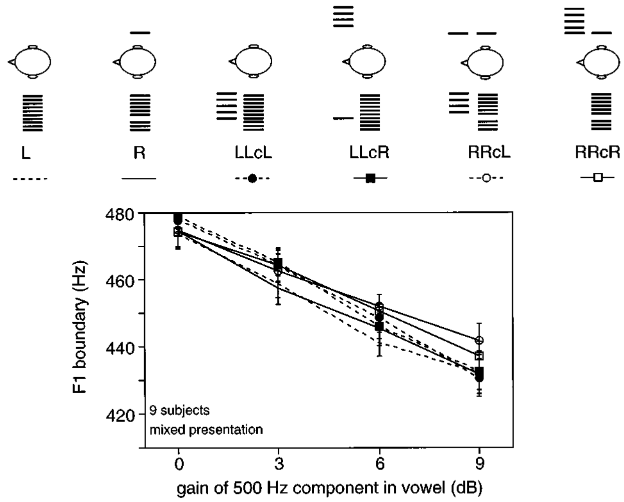

FIG. 9. Phoneme boundaries in experiment 5. The experiment was similar to experiment 4 but used a complex tone as the cue.

the same or different ITDs. The additional complex consisted of the fourth through seventh harmonics of 166.7 Hz. The amplitude of the 500-Hz cue tone which constituted the third harmonic of the complex's harmonic series was the same as in the previous experiment (i.e., identical to that of the 500-Hz harmonic that it preceded), but the amplitude of the higher harmonics fell at -6 dB/oct. The various conditions are shown diagrammatically in Fig. 9. The 500-Hz cue tone always had the same ITD as the 500-Hz harmonic of the vowel. This ITD could be either ± 666 μs. The ITD of the additional complex was also ± 666 μs but varied independently to give four different conditions.

B. Results and discussion

Phoneme boundaries for the two uncued and the four cued conditions are shown in Fig. 9. There are no differences in phoneme boundary across the different conditions ($F_{5,40} = 2.4$, $p > 0.1$). Adding a complex that is harmonically related to the cue tone has prevented it from producing any segregation of the 500-Hz harmonic from the vowel. The segregation found in the first experiment is thus unlikely to be due to processes that are local to 500 Hz. Rather, it is more likely to be due to mechanisms of perceptual organisation which, in the cued conditions of experiment 1, recognize the 500-Hz cue tone as a separate sound source with a distinct location. This source is able to capture the 500-Hz harmonic of the vowel, by the principle of frequency proximity. But experiment 5 has shown that the 500-Hz cue tone loses this ability when it is itself part of a separate harmonic complex.

Capture of the 500-Hz cue tone by the complex occurs regardless of their relative ITDs. For example, there is no significant difference between the uncued L condition and the LLcR condition where, although the cue tone has the same ITD as the complete vowel, the additional complex has the opposite ITD. Here the shared onset, offset, and harmonic relations between the cue tone and the complex override the difference in ITD. This result is another example of the weakness of the grouping exerted by differences in ITD in simultaneous sounds.

It is also clear that there is no across-trial facilitation of segregation by ITD in this experiment. As found in experiment 3, this across-trial facilitation cannot be induced simply by having interspersed trials that have any sound presented with a different ITD. The present experiment has also shown that it is not sufficient to have the cue tone physically present with the appropriate ITD if it is being perceptually integrated into a different percept.

VI. SUMMARY

The present experiments have confirmed previous results showing that in blocks of trials where only the harmonics of a vowel are all present simultaneously, there is no significant segregation of a single harmonic by ITD (experiments 1, 2, and 4).

They have also clarified the conditions under which some perceptual segregation by ITD can occur for simultaneous sounds. Some segregation by ITD (on uncued trials) does occur if other trials are present in the same block in which multiple cue tones (experiment 1) or a single cue tone (experiment 4) are present before the vowel, sharing the harmonic's frequency *and* ITD (experiments 2 and 3). This across-trial facilitation of segregation by ITD is abolished if the cue tone itself forms part of a separate complex tone (experiment 5).

The experiments have also clarified the conditions under which precursor events can segregate a harmonic from a vowel:

(i) a single 500-Hz cue tone is more effective at removing the 500-Hz harmonic from a vowel when the cue tone has a different ITD from the vowel than when it has the same ITD. However, this segregation is not influenced by the ITD of the harmonic in the vowel (experiment 4);

(ii) a single 500-Hz cue tone is *not* effective at removing the 500-Hz harmonic from a vowel when the cue tone is itself captured in a simultaneous harmonic complex (experiment 5). The extent of this capture is itself unaffected by the relative ITDs of the cue and captor tones.

There are two overall conclusions from these experiments. The first confirms Culling and Summerfield's contention that grouping of purely simultaneous sounds by ITD is weak. An individual harmonic cannot be segregated from a vowel on the basis of a difference in ITD unless the listener already has clear evidence for a separate sound source. This evidence must be quite specific: The sound source that induces across-trial segregation by ITD needs to share the ITD and specific tonal quality of the to-be-segregated harmonic; neither narrow-band noise with an appropriate ITD, nor a 500-Hz tone with an inappropriate ITD suffice. The weakness of grouping by simultaneous ITD, when the listener has no such additional evidence for an extra sound source, sits well with experiments that have demonstrated that listeners can integrate ITD across simultaneously present frequencies that form a single perceptual group (Trahiotis and Stern, 1989; Hill and Darwin, 1996). Such across-frequency integration of ITD may help to maintain a stable lateral position for auditory objects in environments where ITDs are disturbed by echoes and reverberation (Woods and Colburn, 1992). However, the present experiments have shown that the auditory system does have the flexibility to perform seg~¡regation by ITD when there is independent evidence of an appropriate additional sound source.

The second conclusion is that sequential grouping by frequency proximity dominates ITD when multiple sounds are present. The single cue tone in experiment 4 was more effective at segregating the harmonic when the cue tone had a different ITD from the vowel; but this effect was barely influenced by the ITD of the harmonic itself. It is likely that this dominance of frequency proximity over ITD acts only over a short timespan, whereas the across-trial facilitation discussed in the previous paragraph is longer term.

An important restriction on grouping by frequency proximity is that it is itself influenced by perceptual grouping. When the cue tone is itself captured, it is ineffective at segregating out the harmonic from the vowel. This restriction has obvious practical advantages, preventing inappropriate segregation when there is an accidental coincidence of frequencies in adjacent sounds. That such capturing itself does not depend on the relative ITDs of the cue tone and its captor harmonics confirms the first conclusion: Grouping of purely simultaneous sounds by ITD is weak.

ACKNOWLEDGMENTS

This research was supported by MRC Grant No. G9505738N. Comments from Dr. Peter Bailey and an anonymous reviewer improved the paper.

[1]The rescaled tanh function $1/1+e^{-s(a-x)}$, where a is the boundary and s the slope parameter provided a sufficiently good fit and allowed convenient boundary estimation using the general curve fit function of the ''Kaleidagraph v.3.0.1'' program (Abelbeck Software) with a and s as free parameters. All curve fits were checked by eye.

Bregman, A. S., and Campbell, J. (**1971**). ''Primary auditory stream segregation and perception of order in rapid sequences of tones,'' J. Exp. Psychol. **89**, 244–249.

Bregman, A. S., and Pinker, S. (**1978**). ''Auditory streaming and the building of timbre,'' Can. J. Psychol. **32**, 19–31.

Cherry, E. C., and Taylor, W. K. (**1954**). ''Some further experiments upon the recognition of speech, with one and with two ears,'' J. Acoust. Soc. Am. **26**, 554–559.

Ciocca, V., and Darwin, C. J. (**1993**). ''Effects of onset asynchrony on pitch perception: adaptation or grouping?,'' J. Acoust. Soc. Am. **93**, 2870–2878.

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., and Gardner, R. B. (**1985**). ''Which harmonics contribute to the estimation of the first formant?,'' Speech Commun. **4**, 231–5.

Darwin, C. J., and Sutherland, N. S. (**1984**). ''Grouping frequency components of vowels: When is a harmonic not a harmonic?,'' Q. J. Exp. Psychol. **36A**, 193–208.

Darwin, C. J., Hukin, R. W., and Al-Khatib, B. Y. (**1995**). ''Grouping in pitch perception: evidence for sequential constraints,'' J. Acoust. Soc. Am. **98**, 880–885.

Darwin, C. J., Pattison, H., and Gardner, R. B. (**1989**). ''Vowel quality changes produced by surrounding tone sequences,'' Percept. Psychophys. **45**, 333–342.

Deutsch, D. (**1979**). ''Binaural integration of melodic patterns,'' Percept. Psychophys. **25**, 399–405.

Deutsch, D. (**1982**). ''Grouping mechanisms in music,'' in *The Psychology of Music*, edited by D. Deutsch (Academic, New York), pp. 99–134.

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.

Russell, P., and Darwin, C. J. (**1991**). ''Real-time synthesis of complex sounds on a Mac II with 56001 DSP chip,'' Br. J. Audiol. **25**, 59–60.

Schubert, E. D., and Parker, C. D. (**1956**). ''Addition to Cherry's findings on switching speech between the two ears,'' J. Acoust. Soc. Am. **27**, 792–794.

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.

van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' Eindhoven University of Technology, Ph.D. dissertation.

Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.

Woods, W. A., and Colburn, S. (**1992**). ''Test of a model of auditory object formation using intensity and interaural time difference discriminations,'' J. Acoust. Soc. Am. **91**, 2894–2902.