Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel

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

(Received 7 September 1994; revised 6 March 1995; accepted 7 April 1995)

The extent to which the 500-Hz component of a steady-state vowel contributed to its phonemic category was measured by estimating the position of the /I/–/ε/ phoneme boundary along a first formant ($F1$) continuum. Shifts in the phoneme boundary were calibrated against shifts produced by physical changes in the level of the 500-Hz component. When the 500-Hz component was presented contralateral to the rest of the vowel, the phoneme boundary changed by an amount equivalent to a physical reduction in the level of the component of about 5 dB (experiment 1). However, giving the 500-Hz component an interaural time difference (ITD) of $-666\ \mu s$ and the remainder of the vowel an ITD of $+666\ \mu s$ did not reduce the component's contribution to vowel quality (experiment 2). Placing the 500-Hz component of the vowel in a sequence of 500-Hz tones substantially reduced the contribution of the 500-Hz component; the contribution was further reduced by giving both the tone sequence and the component an ITD of $-666\ \mu s$ and the rest of the vowel one of $+666\ \mu s$ (experiment 2). When these tone-sequence conditions were presented in the same experimental block as the ITD conditions without a preceding tone sequence, these latter conditions did then show an effect of grouping by ITD. The results suggest that listeners can perceptually segregate sound on the basis of different ITDs, but that this segregation is substantially enhanced if the direction has previously been cued. © *1995 Acoustical Society of America.*

PACS numbers: 43.66.Mk, 43.66.Pn, 43.66.Qp, 43.71.Es

INTRODUCTION

Since different sound sources usually reach a listener from different directions, there is an *a priori* case for the auditory system using directional information to help it to group separately components that originate from different sound sources. Perceptual grouping, though, is only one of at least three possible factors which may be responsible for the increase in intelligibility generally found when a target signal (such as speech) arrives at the listener from a different direction from an interfering signal. Release from monaural (Koenig, 1950; Plomp, 1976) or binaural (Licklider, 1948; Bronkhorst and Plomp, 1988) masking may also contribute.

There is evidence that such perceptual segregation does play a role in both sequential and simultaneous organization. Both speech (Cherry and Taylor, 1954; Huggins, 1964) and sequences of tones (Deutsch, 1979) become less identifiable when they are alternated between the ears. Simultaneously presented components can also be segregated by ear: Listeners can detect much smaller intensity increments in a profile analysis task, (using logarithmically spaced frequency components) when the target component is presented to the same ear as the background than when it is presented to the opposite ear (Green and Kidd, 1983).

Conversely, a number of different studies have found either no, or only weak evidence for perceptual segregation according to ear. Adding noise or another tone to the opposite ear improves the identification of speech or tonal sequences alternated between the ears (Schubert and Parker, 1956; Deutsch, 1979); Deutsch's (1975) scale illusion demonstrates more dramatically the inability, under similar conditions, of the auditory system to organize the input by ear. In simultaneous organization, speech sounds can be readily identified when their component formants are led to opposite ears (Broadbent and Ladefoged, 1957; Rand, 1974; Cutting, 1976; Darwin, 1981), and the pitch of a complex tone can be computed almost as well when its frequency components are distributed across the ears as when its components are played to a single ear (Houtsma and Goldstein, 1972; Beerends and Houtsma, 1989; Darwin and Ciocca, 1992).

The question of whether simultaneous sound can be differentially grouped according to common interaural timing cues has been the subject of recent interest. Culling and Summerfield (1995) presented listeners with two simultaneous "whispered" vowels each of whose first two formants were represented by a pair of noise bands. They found that listeners were no better at identifying the vowel they heard on the left when the noise bands of the target vowel had a different interaural time difference (ITD) ($+390\ \mu s$) from those of the other vowel ($-390\ \mu s$) than when all four noise bands had the same ITD with the left ear leading ($+390\ \mu s$). This result argues against listeners being able to use a common ITD across different frequencies to perceptually segregate simultaneous complex sounds.

Conversely, there is evidence that listeners may be able to use ITDs to help segregate simultaneous sounds. The improvement in identification of pairs of simultaneous vowels that occurs when the two vowels are played on different pitches (Scheffers, 1983) is greater when the vowels also differ in ITD ($\pm 400\ \mu s$) than when they do not (Shackleton and Meddis, 1992; Shackleton *et al.*, 1994). However, it is not clear what are the relative contributions of masking and grouping to this effect.

Evidence that the auditory system can at least *recognize*

0001-4966/95/98(3)/1380/8/$6.00

common ITDs across frequency comes from observations on the lateralization of noise bands (Jeffress, 1972). As a noise band, centered on 500 Hz and with an ITD of 1500 μs, increase in bandwidth from 50 to 200 Hz, it appears to move from the lagging to the leading side (Trahiotis and Stern, 1989). This phenomena can be explained by assuming that the auditory system recognizes common ITDs across frequency either by summing them (Schackleton *et al.*, 1992), or by a more elaborate calculation (Trahiotis and Stern, 1989). A similar phenomenon can be demonstrated for variable depth FM tones (Hukin and Darwin, 1994) and also for different bandwidth harmonic complex sounds (Hill and Darwin, 1993; Hill, 1994).

The present paper examines to what extent the auditory system can use binaural cues to segregate a low-numbered harmonic from the percept of vowel quality. In particular, it asks whether a harmonic that is played to the ear opposite to the remainder of a vowel can still contribute to that vowel's quality (experiment 1); it goes on to ask whether a difference in interaural timing can produce similar effects (experiment 2). The particular paradigm that we use—identification of vowels differing in their first formant frequency, requires the subject to compute the relative levels of harmonics in the first formant region. It thus has some similarities with the profile analysis task referred to above, where clear segregation occurred by ear. Conversely, the speech and pitch tasks described above, that have shown little evidence for segregation by ear, do not critically depend on the relative levels of different components across the ears. The paradigm has the advantage over experiments requiring the identification of pairs of simultaneous vowels in that all the frequency components relevant to the calculation of the first formant frequency are well resolved by the auditory system and so should be little affected by monaural or binaural masking effects; they are also in the frequency range in which ITD is an effective cue to lateralization. The identification of a single vowel is also a simpler task than the identification of paris of vowels, where the mechanisms underlying the improvement in identification with differences in fundamental frequency are complex and not fully understood (Culling and Darwin, 1993; Culling and Darwin, 1994).

In the following experiments we are concerned solely with the extent to which lateralization cues can segregate a component from a harmonic vowel for the purpose of identifying its vowel quality. This is logically a separate question from whether the 500-Hz component can be heard out as a separate sound, or whether it forms a distinct auditory stream. There is considerable evidence that the effectiveness of a cue in segregating a sound may differ substantially depending on how "segregation" is measured (Darwin and Carlyon, 1995; Hukin and Darwin, 1995).

I. EXPERIMENT 1

The first experiment establishes a paradigm for investigating the extent to which binaural information influences the contribution that a single harmonic makes to vowel quality. The paradigm is similar to one used previously to investigate the effects of onset asynchrony (Darwin, 1984) and inharmonicity (Darwin and Gardner, 1986) on segregating a single harmonic from a vowel percept, but has been extended to allow a quantitative calibration of the effect against actual changes in the level of the harmonic in question.

The paradigm relies on the fact that the distinction between /I/ 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 /I/ or /ɛ/; their phoneme boundary is then established at the 50% identification point. In our previous experiments this boundary normally occurs at an $F1$ at around 450 Hz. The 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 /I/–/ɛ/ phoneme boundary. The technique is sensitive to changes of around 1 dB in the level of harmonics close to $F1$. In this experiment, we first calibrate the /I/–/ɛ/ phoneme boundary against changes in level of the 500-Hz component of the vowel; we then ask how the contribution of this harmonic to the calculation of $F1$ is changed by presenting it either diotically or dichotically to the monotic remainder of the vowel. If the 500-Hz tone makes a reduced contribution to the vowel quality as a result of these manipulations, then the phoneme boundary should shift to higher nominal $F1$ values.

A. Stimuli

The vowels presented to listeners differed along three dimensions: (i) The $F1$ frequency used to synthesize them; (ii) the additional gain given to the 500-Hz component; and (iii) whether presentation of the 500-Hz component was ipsilateral to the rest of the vowel, contralateral to the rest of the vowel or diotic. The six different $F1$'s varied between 395 and 500 Hz in 21-Hz steps to give a percept that changed from /I/ at low values of $F1$ to /ɛ/ at high values. With ipsilateral presentation seven additional gains were given to the 500-Hz component: -9, -6, -3, 0, $+3$, $+6$, or $+9$ dB. With diotic and contralateral presentation only the last four gains were used. The actual level of the 500-Hz tone in a particular stimulus was thus a function of both the $F1$ value and the additional gain. In the $+0$-dB gain condition the level of the 500-Hz component was 10.5 dB higher when $F1$ was 500 Hz then when $F1$ was 395 Hz.

The remaining, constant stimulus properties were as follows. The second and third formant frequencies were fixed at 2100 and 2900 Hz, respectively. The half-power 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 had a duration of 56 ms. The vowel was always presented to the left ear.

The sounds were digitally synthesized in real time with a sample rate of 44.1 kHz using custom software (Russell, 1992) on a Digidesign Audiomedia board attached to an Apple Mac IIfx which also controlled the experiment. The vowel spectra were produced by summing harmonics with amplitudes and phases given by the serial vocal tract transfer function of Klatt (1980). The sounds were output through the Audiomedia's 16-bit DACs and antialiasing filters, low-pass filtered at 10 kHz and amplified using a custom-built headphone amplifier. Subjects heard the stimuli over Sennheiser HD414 headphones in a double-skinned IAC booth; the 500-Hz component of the member of the original continuum with an $F1$ of 500 Hz was at a level of 60 dB SPL.

B. Procedure

The experiment was run in two blocks: One block used only ipsilateral presentation and the other used both diotic and contralateral. The ipsilateral block contained 420 trials (ten replications of six $F1$ values at each of the seven gains) and the diotic and contralateral block contained 480 trials (ten replications of six $F1$ values at each of the four levels for two lateral positions). The trials in each block were presented in different, pseudorandom orders. 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 their right ear, which they were to ignore. They signaled their response on each trial using the "i" and "e" keys on the Mac keyboard. Each sound followed 500 ms after the response to the previous one. Subjects could repeat the previous sound by pressing the "escape" key. Each experimental block was immediately preceded by a short familiarization session, in which the subjects were given one example of each stimulus used in that block.

Nine subjects took part in the experiment, all of whom had previous experience of 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.

C. Results

Phoneme boundaries for each subject were estimated from the number of "i" responses to the ten repetitions of the six stimuli, differing in $F1$, in each condition. The $F1$ boundary for each subject was estimated by a least-squares fit of a rescaled tanh function[1] to the data in each of the 15 different gain and presentation conditions.

The average $F1$ boundaries across subjects are shown in Fig. 1. Altering the gain of the 500-Hz component has a clear effect on the phoneme boundary with ipsilateral, diotic, and contralateral presentation. All three show an almost linear relationship (accounting for at least 98% of the variance) between the change in gain of the 500-Hz component (in dB) and the $F1$ boundary. As the gain of the 500-Hz component is increased above its original value (0 dB), the phoneme boundary moves to lower $F1$ values. This decrease in the boundary frequency is presumably due to the perceptually estimated $F1$ value increasing with the increased level of the harmonic just above the formant frequency. A similar, but opposite change occurs as the 500-Hz component's gain is reduced below its original value.

Regression lines fitted through each set of data point have slopes (and standard errors of fit) of −5.9 Hz/dB (1.2 Hz), −4.8 Hz/dB (0.88 Hz), and −3.6 Hz/dB (0.18 Hz) for ipsilateral, diotic, and contralateral presentation, respectively. Analysis of variance on the 0- to +9-dB data points showed significant main effects of condition ($F_{2,16}=51.4$, $p<0.0001$), and of level ($F_{3,16}=74.6$, $p<0.0001$) and a significant interaction between conditions and level ($F_{6,48}=4.1$, $p<0.01$). These significant effects were due almost entirely to contralateral presentation being different from ipsilateral ($F_{1,8}=347.0$, $p<0.0001$) and diotic ($F_{1,8}=260.4$, $p<0.0001$), which barely differed ($F_{1,8}=6.2$, $p<0.02$).

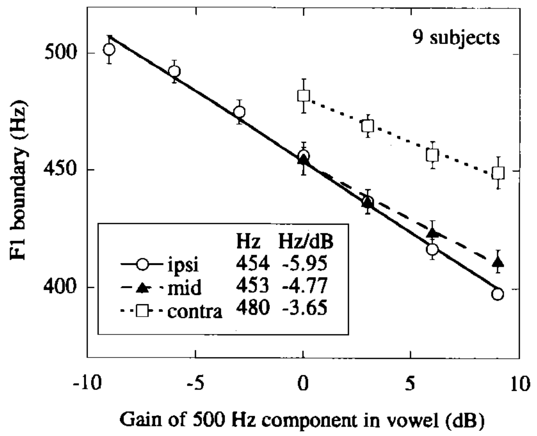

FIG. 1. First formant phoneme boundaries between /ɪ/ and /ɛ/ and standard errors across nine subjects for 56-ms steady-state vowels in experiment 1 as a function of the gain of the vowel's fourth harmonic (500 Hz) for conditions where that component is presented either ipsilateral (ipsi) or contralateral (contra) to the remainder of the vowel, or diotically (mid).

Looking first at the difference between the ipsilateral and contralateral presentation: Putting the 500-Hz component at its original level (gain of 0 dB) into the contralateral rather than ipsilateral ear leads to an increase in phoneme boundary frequency of about 30 Hz, which is equivalent to a reduction in gain of the 500-Hz component of about 5 dB. The direction of this change is appropriate for an explanation in terms of auditory grouping, but not with one based on release from ipsilateral masking (Rand, 1974) or a change in loudness mediated by suppression (Moore *et al.*, 1985). If the effect of putting the 500-Hz component into the opposite ear were simply to make it more prominent or louder, the boundary change should have been in the opposite direction. But if the effect of putting the 500-Hz component into the opposite ear were to remove it (partially) from the calculation of vowel quality, then we would expect a shift in phoneme boundary in the direction that we have found. It is possible though that this shift may have been reduced by the 500-Hz tone being heard as louder in the contralateral than in the ipsilateral condition.

The data from contralateral presentation also have a shallower slope than the ipsilateral data, so that increases in gain have less effect on the phoneme boundary with contralateral presentation than with ipsilateral. It is not clear what causes this difference. It is possible that partial perceptual removal of a component is not equivalent to a simple reduction in level, so that the apparent reduction is level dependent. Alternatively, it may be that a component that exceeds by too much its normal level in a vowel will be more effectively grouped out of the vowel than components that plausibly fall under a vowel-like envelope (Darwin, 1984).

The lack of any substantial difference between ipsilateral and diotic presentation is harder to interpret because of possible changes in loudness of the 500-Hz tone caused by it being presented to both ears. For instance, the slight upward shift in phoneme boundary could the result of a more substantial upward shift caused by grouping out the 500-Hz component being reduced by a loudness increase due to binaural loudness summation.

D. Discussion

The first experiment has shown that presenting the 500-Hz component of a vowel to the opposite ear from the remainder of the vowel produces a phoneme-boundary shift equivalent to a physical reduction of about 5 dB in its level. Since the loudness of the 500-Hz component may have increased by presenting it dichotically rather than monotically to the rest of the vowel, the 5-dB reduction might have been greater had we matched loudness rather than objective gain across ipsilateral and contralateral presentation. Similarly, presenting the 500-Hz component diotically may have increased its loudness relative to the other two conditions thus increasing its contribution to the vowel. Whatever the relative role of grouping and changes in loudness to the effects that this experiment has found, the substantial difference in phoneme boundary between ipsilateral and contralateral presentation has provided a tool to investigate the effectiveness of interaural time differences (ITDs) for auditory grouping in vowel perception.

Although contralateral presentation of the 500-Hz component does reduce its effective contribution to the quality of the vowel, it is clear from the slope of the contralateral function that changes in its level are still influencing the vowel's quality. Subjects can apparently compute a spectral envelope in the $F1$ region from the levels of components that go to different ears. This result contrasts with the difficulty that subjects have with contralateral presentation of a single component in a profile analysis task using logarithmically spaced components (Green and Kidd, 1983). It is possible that this difference between ipsilateral and contralateral presentation would be reduced if the profile analysis task were done with harmonically related rather than logarithmically spaced components, but unpublished experiments by Kidd (personal communication) have not shown any substantial improvement in the dichotic profile analysis task when harmonically related sounds are used.

II. EXPERIMENT 2

The second experiment asked whether listeners can segregate one harmonic from a vowel using ITDs rather than the dichotic versus monotic presentation used in the first experiment. A pilot experiment had indicated that ITDs gave rather weaker effects than those in the first experiment, so experiment 2 also included another manipulation designed to enhance the grouping effect of ITDs.

The idea behind the additional manipulation is that the auditory system may be more able to segregate one component of a simultaneous complex differentially if it has recently heard a sound coming from the direction of that component. For example, there is evidence from experiments on dichotic listening to consonant–vowel syllables that selective attention to one ear is more effective if the to-be-attended ear is cued by a brief tone, than if the subject is merely instructed to report the syllable on that ear (Mondor and Bryden, 1991; Mondor and Bryden, 1992). A cue tone on one side can also speed-up decisions made about the location of sounds subsequently presented to that side (Spence and Driver, 1994). Grouping mechanisms also need to organize the auditory input into potential sound sources that may be extended in time. A powerful sequential grouping principle is frequency proximity, whereby tones that are similar in frequency tend to be grouped together into a melodic line (Bregman and Campbell, 1971; Dowling, 1973). This principle can be used to segregate energy added to a single harmonic of a vowel. Using the same /I/–/ɛ/ identification task as used in experiment 1, Darwin *et al.* (1989) showed that the phoneme boundary shift produced by giving the 500-Hz component of each vowel 6 dB added energy is at least partly removed from the calculation of vowel quality if the vowel is preceded by a train of four tones at 500 Hz. This removal probably occurs because the added 500-Hz energy forms a separate auditory stream with the preceding tone sequence. The removal of this extra energy from the vowel at the harmonic frequency is more complete when the vowel is also followed by two similar tones.

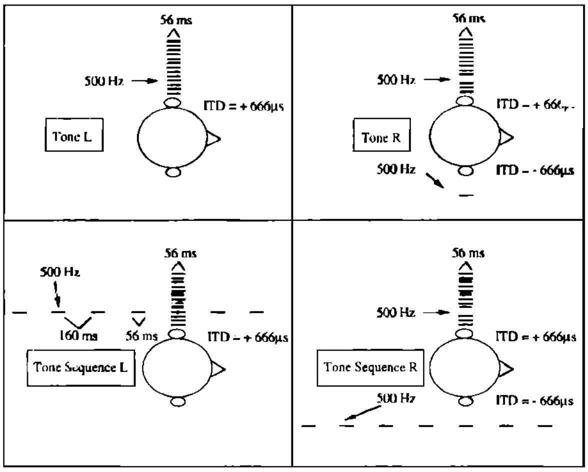

FIG. 2. Schematic of sounds used in experiment 2. All the sounds are presented binaurally, but with different ITDs. The bottom right panel illustrates the tone sequence R condition in which the vowel (minus its 500-Hz component) is presented with an ITD of $+666$ μs, whereas the sequence of 500-Hz tones (including the 500-Hz component of the vowel) is presented with an ITD of -666 μs.

Experiment 2 uses a similar manipulation to try to obtain evidence for grouping by ITD. It includes conditions in which the vowel is embedded in a sequence of tones at 500 Hz, and asks whether the 500-Hz component is removed more effectively from the vowel when both it and the sequence of tones have a different ITD from the rest of the vowel than when they have the same ITD. Figure 2 shows the stimulus configuration for the condition in which the 500-Hz tone sequence (including the 500-Hz tone that is nominally part of the vowel) is presented with an ITD that localizes it to the ear opposite to the rest of the vowel.

The experiment also looks at whether any cuing effect of the sequence of tones is specific to a particular trial or whether the presence of a tone sequence on one trial may help segregation by ITD on a subsequent trial that lacks the tone sequence. In order to test whether such across-trial effects occur, the experiment compares blocked and mixed presentation. In the blocked conditions, trails with the tone sequence are presented in a different block from those without the tone sequence. In the mixed condition, the different types of trials are all randomized together. If across-trial effects are taking place then we would expect to find increased evidence for segregation by ITD in the mixed over the blocked case for trials on which the tone sequence is absent.

To summarize: This experiment asks three questions. First, can an ITD perceptually segregate a single harmonic from a vowel? Second, is this segregation enhanced by an appropriate preceding cue? Third, is the effect of this cue restricted to a particular trial?

A. Stimuli and procedure

The 0-, +3-, +6-, and +9-dB gain manipulations from experiment 1 were used here but in four different presentation conditions. In all conditions the vowel components other than the 500-Hz were all presented binaurally with an ITD that led on the left ear by 666 μs. In the "tone L" condition the 500-Hz tone was also presented with the same ITD. In the "tone R" condition the ITD for the 500-Hz tone led on the right ear by 666 μs. These stimulus conditions are illustrated in the upper part of Fig. 2.

Two other conditions embedded the vowel in a brief sequence of 500-Hz tones. In the "tone sequence L" condition the vowel was preceded by four and followed by two 56-ms, 500-Hz tones separated by 160 ms. The level of the 500-ms tones in the sequence was the same as the level of the 500-ms tone in the vowel in that particular condition (this level varied as before depending on the $F1$ of the vowel and the additional gain that had been given it). All the sounds had an ITD of 666 μs with the left ear leading.

The "tone sequence R" condition was similar except that all the 500-Hz tones (including that in the vowel) had an ITD of 666 μs with the right ear leading. ITDs were produced by delaying the entire waveform to one ear, so that both the fine structure of the sound and the envelope were delayed.

Each of these four conditions was presented to seven listeners (six of whom had taken part in experiment 1) in two ways. With blocked presentation, the same subjects heard the tone L and tone R conditions in one block and the tone sequence L and tone sequence R conditions in a different block; block order was counterbalanced across subjects. With mixed presentation subjects heard all four conditions randomized together. The procedure was otherwise similar to experiment 1.

B. Results

1. Blocked presentation

Mean phoneme boundaries across the seven subjects for the tone L and tone R conditions under blocked presentation are given by the open symbols in Fig. 3. The increase of gain of the 500-Hz component progressively lowers the phoneme boundary as in experiment 1 ($F_{1,6}=15.6$, $p<0.01$). As before, the change is linear with slopes (standard error of fit) of −4.8 Hz/dB (0.51 Hz) and −4.5 Hz/dB (0.39) for tone L and tone R, respectively. In the tone R condition the tone is still being (partially) incorporated into the calculation of vowel quality. The two conditions do however differ significantly overall ($F_{1,6}=15.6$, $p<0.01$), with the phoneme boundaries in the tone R condition being at a higher frequency than in the tone L condition. The difference in ITD between the 500-Hz tone and the rest of the vowel in the tone R condition therefore does allow some perceptual segregation from the vowel. However, the size of this effect is small, amounting to a shift of only 10 Hz in $F1$ (equivalent to −2 dB) at 9 dB compared with a shift of some 50 Hz (equivalent to −8.4 dB) from ipsilateral to contralateral at 9-dB gain in experiment 1. ITD thus provides much weaker segregation for a single harmonic than does the contralateral presentation of experiment 1.

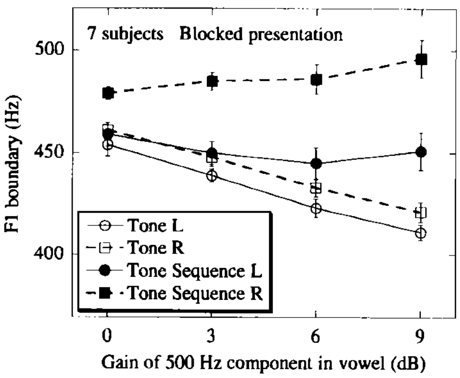

FIG. 3. First formant phoneme boundaries between /I/ and /ε/ and standard errors across seven subjects for 56-ms steady-state vowels in experiment 2 as a function of the gain of the vowel's fourth harmonic (500 Hz). The four different stimulus conditions were presented in separate blocks of trials.

When the 500-Hz tone is embedded in a sequence of similar tones, the results are quite different as shown by the filled symbols in Fig. 3. There are two differences. First, the tone sequence removes the effect of the gain of the 500-Hz component ($F_{1,6}=1.4$, $p>0.1$), giving essentially flat functions in Fig. 3; second there is a substantial effect of ITD, with phoneme boundaries shifting to higher $F1$ values with contralateral presentation when the 500-Hz tones have a different ITD to that of the rest of the vowel ($F_{1,6}=15.3$, $p<0.1$). The slopes (standard error of fit) for the tone sequence L and R conditions are respectively −1.0 dB/Hz (2.21 Hz) and 1.7 (1.06).

The effect of embedding the 500-Hz tone in a sequence of similar tones confirms previous results that a surrounding tone sequence can perceptually remove energy from a vowel at the tones' frequency (Darwin *et al.*, 1989). The ipsilateral tone sequence (tone sequence L) gives boundary shifts equivalent to removing rather little energy in the 0-dB gain condition (compared with the tone L condition), but equivalent to significant amounts in the 6- and 9-dB conditions.

Unfortunately this experiment did not include a condition in which the 500-Hz component was physically removed from the vowel. This condition would have provided a baseline against which to measure the effectiveness of different conditions for perceptually removing the 500-Hz com-

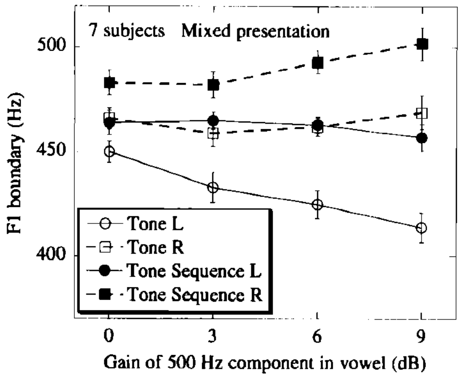

FIG. 4. Results in experiment 2 as in Fig. 3, but the four different stimulus conditions were presented randomized together in the same block of trials.

ponent. However, the effect of physical removal found in each of two previous experiments that have used very similar stimuli was to give an increase in the phoneme boundary of about 25 Hz (Darwin and Gardner, 1986; Hukin and Darwin, 1995). A somewhat larger change would probably have been found in this experiment, since a gain change in the 500-Hz component of −9 dB in experiment 1 gave an increase of 45 Hz in the phoneme boundary. An increase of 33 Hz is found in the tone sequence R condition at 0 dB, rising to 52 Hz for the +9-dB condition (relative to tone L 0 dB) indicating that a contralateral 500-Hz tone sequence removes a substantial part of the energy at the frequency from the vowel.

2. *Mixed presentation*

Under mixed presentation the results (Fig. 4) are very similar to those with blocked presentation, but with one important exception: There is now a substantial difference between the tone L and the tone R conditions. Increasing the gain of the 500-Hz component in the tone R condition now has almost no effect on the $F1$ boundary (slope=0.4 Hz/dB; standard error=1.0 Hz), indicating that ITD is effective even without a surrounding tone sequence at perceptually removing at least part of the 500-Hz component from the vowel. The tone R condition behaves very similarly with mixed presentation as the tone sequence L condition. In an analysis of variance on the tone L and tone R conditions, there is a significant main effect of tone L/tone R presentation ($F_{1,6}=39.7$, $p<0.001$) and a significant interaction between presentation and gain ($F_{1,6}=6.8$, $p<0.005$).

C. Discussion

This experiment shows, first, that a large ITD difference by itself produces less segregation than the contralateral presentation used in experiment 1. A difference of ±666 μs in ITD between the 500-Hz component of a vowel and the remaining components produces a much smaller reduction in the contribution made by that component to the vowel for the calculation of vowel quality than does presenting the 500-Hz component at the opposite ear. This very weak effect of ITD on grouping for vowel identification is compatible with the failure of Culling and Summerfield (1995) to find any evidence for differential grouping of the formants of a vowel pair by ITD. The difference that we have found between contralateral and ITD presentation would not have been predicted from the finding (Shackleton *et al.*, 1994) that contralateral presentation was no more effective than a 400-μs ITD at improving the identification of simultaneous vowel pairs, but it is possible that masking rather than grouping effects were responsible for the improvement (Culling and Summerfield, 1995).

Second, this experiment has also shown that the weak segregation produced by an ITD for isolated simultaneous sounds can be substantially increased by the temporal context of a series of similar tone. If the tone-to-be-segregated is part of a brief sequence of tones which all share the same ITD, then there is substantial segregation from a vowel which has a different ITD. A consistent feature of the data from the two experiments is that the slopes of the functions relating $F1$ boundary to gain of the 500-Hz component are flatter in condition where there is more segregation of the 500-Hz component from the vowel. A completely flat function indicates that the effective contribution of the 500-Hz component to the vowel is the same, regardless of its actual level. This result is not surprising in the sequence conditions where the surrounding tone sequence has the same level as the 500-Hz component—a more intense surrounding might be expected to pull out more either by adaptation or by grouping. The tone sequence L condition of experiment 2 however contrasts with the condition with contralateral presentation in experiment 1. The former gives a flat function whereas the latter gives one with a significant, albeit reduced, slope. It is not clear why these two conditions should give different slopes, nor indeed why ITDs should be less effective than ILDs in segregating a harmonic from a vowel. Further experiments are needed to examine parametrically the effect of different sizes of ITD and ILD and the relation of their segregating effect to subjective position of the 500-Hz component, both in isolation and when presented with the vowel. Our informal impression is that contralateral presentation is more effective at letting the listener hear out the 500-Hz component in a separate spatial position than is an ITD of 666 μs.

Third, the increased segregation produced by this temporal context generalizes to trials in the same block that do not have the context. This generalization across trial suggests that subjects' attention might be being directed to possible sounds on the right side by the obvious presence of a tone series on that side during other trials. If this effect is a general attentional one, then it should persist when spatial direction is cued by sounds that are unrelated to those that are to be segregated. A temporal context consisting of short noise bursts, rather than tone bursts, should produce the across-trial generalization. A further experimental question is whether both the segregation and the generalization would still occur when the lateralization of the tone sequences is cued by intensity differences rather than ITDs.

III. SUMMARY AND CONCLUSIONS

These two experiments have measured the extent to which the 500-Hz, 4th harmonic of a front vowel can be segregated from it by contralateral presentation in interaural time (±666 μs) differences. The amount of segregation was measured by the change in the /ɪ/–/ɛ/ phoneme boundary along an $F1$ continuum.

(i) When the 500-Hz component of vowels along an $F1$ continuum was physically changed in level by up to ±9 dB, the /ɪ/–/ɛ/ phoneme boundary changed linearly by −6 Hz/dB.

(ii) Putting the 500-Hz component (with up to 9 dB increase in level) in the opposite ear from the remainder of the vowel, produced an increase in the phoneme boundary equivalent to a reduction of between 5 and 8 dB in its level (depending on its gain), indicating that the 500-Hz component had been partially segregated from the vowel.

(iii) A much smaller increase in the phoneme boundary resulted when the 500-Hz component was given a different ITD from the remainder of the vowel.

(iv) Embedding the 500-Hz component in a short series of similar tones substantially increased the segregation of the tone from the vowel.

(v) This segregation was further substantially increased when all the 500-Hz tones were given a different ITD from the remainder of the vowel.

(vi) The small effect of ITD described in (iii) above was substantially increased when trials of that condition were mixed with trials of the conditions in (iv) and (v).

ACKNOWLEDGMENTS

This research was supported by Grant G9116710N from the UK MRC to the second author. We are grateful to John Culling and Quentin Summerfield for extensive discussions and comments on an earlier draft and for the comments of two anonymous reviewers.

---

[1] The rescaled tanh function

$$\frac{1}{1+e^{-s(a-x)}},$$

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.

Beerends, J. G., and Houtsma, A. J. M. (**1989**). "Pitch identification of simultaneous diotic and dichotic two-tone complexes," J. Acoust. Soc. Am. **85**, 813–819.

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.

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

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.

Culling, J. E., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by Fo," J. Acoust. Soc. Am. **93**, 3454–3475.

Culling, J. E., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.

Culling, J. F., and Summerfield, Q. (**1995**). "Binaural masking release of complex sounds: Absence of across frequency grouping by common interaural delays," J. Acoust. Soc. Am. **98** (to be published).

Cutting, J. E. (**1976**). "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," Psychol. Rev. **83**, 114–140.

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. (**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).

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

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

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am. **57**, 1156–1160.

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

Dowling, W. J. (**1973**). "The perception of interleaved melodies," Cognitive Psychol. **5**, 322–337.

Green, D. M., and Kidd, G. (**1983**). "Further studies of auditory profile analysis," J. Acoust. Soc. Am. **73**, 1260–1265.

Hill, N. I. (**1994**). "Binaural processing of a multi-sound environment: The role of auditory grouping cues," Sussex, D.Phil.

Hill, N. I., and Darwin, C. J. (**1993**). "Effect of onset asynchrony and of mistuning on the lateralization of a pure tone embedded in a harmonic complex," J. Acoust. Soc. Am. **93**, 2307–2308.

Houtsma, A. J. M., and Goldstein, J. L. (**1972**). "The central origin of the pitch of pure tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. **51**. 520–529.

Huggins, A. W. F. (**1964**). "Distortion of the temporal pattern of speech: Interruption and alternation," J. Acoust. Soc. Am. **36**, 1055–1064.

Hukin, R. W., and Darwin, C. J. (**1994**). "Lateralization of a frequency-modulated 500-Hz tone," J. Acoust. Soc. Am. **95**, 2895–2896.

Hukin, R. W., and Darwin, C. J. (**1995**). "Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification," Percept. Psychophys. **57**, 191–196.

Jeffress, L. A. (**1972**). "Binaural signal detection: Vector theory," in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic, New York).

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

Koenig, W. (**1950**). "Subjective effects in binaural hearing," J. Acoust. Soc. Am. **22**, 61–62.

Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.

Mondor, T. A., and Bryden, M. P. (**1991**). "The influence of attention on the dichotic REA," Neuropsychologia **29**, 1179–1190.

Mondor, T. A., and Bryden, M. P. (**1992**). "Orienting of auditory spatial attention: Effects of a lateralized tone cue," Neuropsychologia **30**, 743–752.

Moore, B. C. J., Glasberg, B. R., Hess, R. F., and Birchall, J. P. (**1985**). "Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears," J. Acoust. Soc. Am. **77**, 1505–1513.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Rand, T. C. (**1974**). "Dichotic release from masking of speech," J. Acoust. Soc. Am. **55**, 678–680.

Russell, P. (**1992**). *Synthesizer: User Manual*, Laboratory of Experimental Psychology, University of Sussex, Brighton, U.K.

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

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.

Shackleton, T. M., and Meddis, R. (**1992**). "The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs," J. Acoust. Soc. Am. **91**, 3579–3581.

Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). "Across frequency integration in a model of lateralization," J. Acoust. Soc. Am. **91**, 2276–2279.

Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1994**). "The role of binaural and fundamental frequency difference cues in the identification of concurrently presented vowels," Q. J. Exp. Psychol. **A 47**, 545–563.

Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," J. Exp. Psychol.: Human Percept. Perf. **20**, 555–574.

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.