Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention

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

(Received 29 June 1999; revised 31 January 2000; accepted 28 March 2000)

Three experiments explored the resistance to simulated reverberation of various cues for selective attention. Listeners decided which of two simultaneous target words belonged to an attended rather than to a simultaneous unattended sentence. Attended and unattended sentences were spatially separated using interaural time differences (ITDs) of 0, ±45, ±91 or ±181 μs. Experiment 1 used sentences resynthesized on a monotone, with sentence pairs having $F0$ differences of 0, 1, 2, or 4 semitones. Listeners' weak preference for the target word with the same monotonous $F0$ as the attended sentence was eliminated by reverberation. Experiment 1 also showed that listeners' ability to use ITD differences was seriously impaired by reverberation although some ability remained for the longest ITD tested. In experiment 2 the sentences were spoken with natural prosody, with sentence stress in different places in the attended and unattended sentences. The overall $F0$ of each sentence was shifted by a constant amount on a log scale to bring the $F0$ trajectories of the target words either closer together or further apart. These prosodic manipulations were generally more resistant to reverberation than were the ITD differences. In experiment 3, adding a large difference in vocal-tract size (±15%) to the prosodic cues produced a high level of performance which was very resistant to reverberation. The experiments show that the natural prosody and vocal-tract size differences between talkers that were used retain their efficacy in helping selective attention under conditions of reverberation better than do interaural time differences. © *2000 Acoustical Society of America.* [S0001-4966(00)02607-2]

PACS numbers: 43.66.Pn, 43.71.Es [RVS]

INTRODUCTION

This paper is concerned with some of the cues that listeners can use to attend to a particular sound source over time. It extends to conditions of reverberation from simulated room acoustics the findings of a recent article (Darwin and Hukin, 1999) and its companion article (Darwin and Hukin, 2000) on the effectiveness of spatial, prosodic, and vocal-tract size cues to auditory selective attention.

Reverberation has a variety of destructive influences on the intelligibility of speech, both for single sound sources (Moncur and Dirks, 1967; Nabelek and Robinson, 1982; Nabelek and Donahue, 1984; Nabelek and Dagenais, 1986; Nabelek, 1988) and when there are competing sounds (Plomp, 1976, 1977; Culling *et al.*, 1994). In this article we examine the effect that reverberation has on some of the cues that can potentially help listeners to attend to a particular talker across time.

Two types of cue that can potentially help a listener to maintain attention to a particular sound source have dominated discussions and were investigated in the companion article: localization and pitch. Both of these cues, however, are susceptible to adulteration by reverberation. Although single sounds with abrupt onsets are well localized in naturally reverberant or simulated reverberant environments due to the mechanism of the precedence effect (Hartmann, 1983; Culling *et al.*, 1994), localization of sounds that lack abrupt onsets is seriously impaired by reverberation (Hartmann, 1983), because of changes to both interaural time and intensity differences (Rakerd and Hartmann, 1985). When masking noise is present, the ability to localize speech (Abouchacra *et al.*, 1998) or click-trains (Good and Gilkey, 1996; Lorenzi *et al.*, 1999) is impaired at adverse signal-to-noise ratios in anechoic conditions; the influence of reverberation on this ability has not been studied systematically. However, even modest amounts of reverberation, which do not reduce listeners' ability to localize speech presented alone, can reduce listeners' ability to exploit localization cues in identifying a vowel target presented with spatially separated masking noise (Culling *et al.*, 1994).

If the $F0$ of a complex sound is steady, it should be little affected by reverberation, since the harmonic structure remains intact. However, the harmonic structure of frequency-modulated sounds is distorted by reverberation since each part of the reverberant sound, being delayed, will have a previous value of $F0$ rather than that of the current direct sound. Figure 1 shows spectrograms of the sentence ''Could you please write the word bead down now'' spoken with natural prosody for both the anechoic (upper panel) and reverberant (lower panel, $\mathrm{RT}_{60}=0.4\,\mathrm{s}$) conditions used in the following experiments. Where the $F0$ contour is relatively flat, harmonic structure is still evident though with reduced clarity, but distortion of harmonicity is clearly visible where there are large changes in $F0$ (during the word ''now,'' for example).

The effect of this degradation has been shown in experiments on the recognition of double vowels (Culling *et al.*, 1994). For vowels with steady $F0$'s, the improved identification produced by putting the vowels on different $F0$'s survives reverberation. But for vowels with modulated $F0$'s it does not. These findings raise the question of whether more

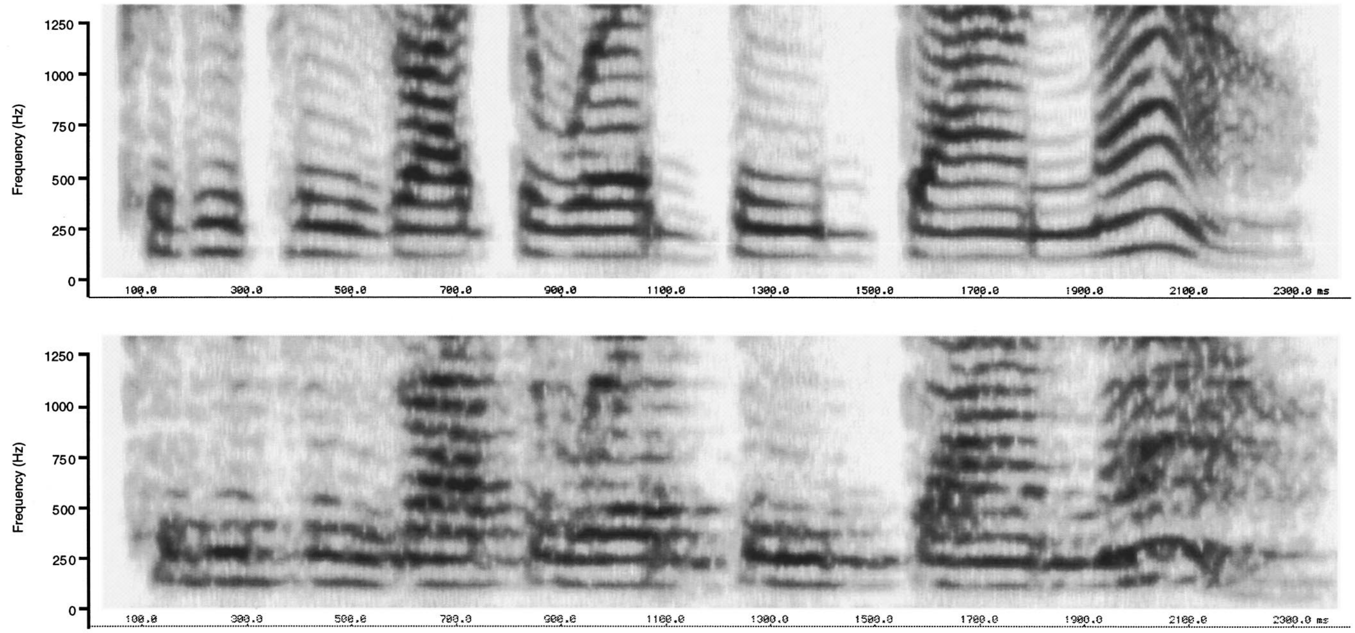

FIG. 1. Narrow-band spectrograms of the sentence ‘‘Could you please write the word bead down now’’ spoken with natural prosody for both anechoic (upper panel) and reverberant (lower panel, $\mathrm{RT}_{60}=0.4$ s) conditions. The reverberation produces considerable smearing across time and distortion of harmonicity where there are rapid changes in $F0$ (during the word ‘‘now,’’ for example).

natural intonation contours are useful under reverberant conditions for selecting between alternative sound sources. Listeners might, for example, be able to attend more easily to an on-going natural contour than to artificial modulation, and so overcome the degrading effects of reverberation.

The experiments reported here use an established paradigm (Darwin and Hukin, 1999, 2000) to investigate the effects of reverberation on cues to speech source continuity. Subjects choose which of two simultaneous target words are part of an attended sentence rather than part of another simultaneous sentence. The paradigm has the advantage that it allows a rapid investigation of the effectiveness of localization and of prosodic and speaker cues in determining speech source continuity, although it does not measure the real-time allocation of attention. Since the two sentences and the target words remain the same throughout the experiment (apart from the manipulated cues), the intelligibility requirements of the task are minimal.

Our paradigm complements recent work by Assmann (1999b, a) which investigates how $F0$ and vocal-tract size differences contribute to the overall intelligibility of pairs of sentences. Assmann measures the total number of words recalled irrespective of which of the two sentences a particular word occurred in. Consequently, his work asks how various cues influence the intelligibility of individual words, but does not address the question of how listeners determine which words are part of the attended sentence. Our paradigm ignores the former question, and addresses the latter.

The first two experiments use a simulated room (Peterson, 1986) to explore the effects of reverberation on the usefulness of localization and prosodic cues to selective attention. The third experiment also varies the apparent vocal-tract size of the talker.

I. EXPERIMENT 1

This experiment repeats experiment 1 of Darwin and Hukin (1999) with simulated reverberation. The experiment examines the robustness to simulated reverberation of monotonous $F0$ differences and interaural time differences as cues for the selection of one of two simultaneous target words.

A. Stimuli

The recordings from the earlier article (Darwin and Hukin, 1999) were used in this experiment. The two sentences ‘‘Could you please write the word bird down now’’ and ‘‘You will also hear the sound dog this time’’ were spoken with a nearly flat intonation contour at around 125 Hz by a native speaker of British English (CJD). A short period of silence was added to the beginning of one sentence so that the two target words (‘‘dog,’’ ‘‘bird’’) began at the same time into their respective sound files.

The two sentences were resynthesised on a monotone using a PSOLA algorithm (Moulines and Charpentier, 1990) at fundamental frequencies of 100, 106, 112.3, and 125 Hz, corresponding to approximately 0, 1, 2, and 4 semitones above 100 Hz. This range of $F0$ differences is sufficient to produce substantial segregation both in speech identification tasks (Brokx and Nooteboom, 1982; Scheffers, 1983; Assmann and Summerfield, 1990; Culling and Darwin, 1993; Bird and Darwin, 1998) and in across-frequency integration of interaural time differences (ITDs) (Hill and Darwin, 1996).

In order to maintain the alignment of target word onsets, small adjustments were made to the silent closure interval before the target word in the different $F0$ conditions. These adjustments compensated for the PSOLA resynthesis round-

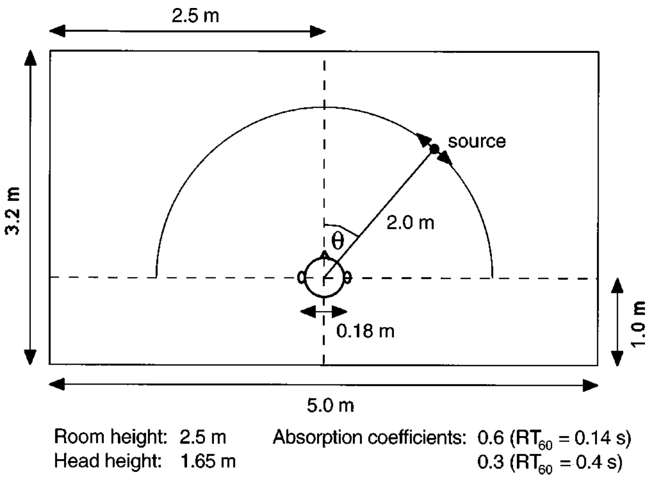

FIG. 2. Layout of simulated room for implementation of Peterson's ray-tracing method. The head was acoustically transparent, thus eliminating head-related interaural intensity differences (apart from those arising from the distance between the ears). Sources were positioned to give the nominal ITDs cited in the text when the absorption coefficient was unity.

ing durations to whole numbers of pitch periods.

This procedure produced sounds for the normal condition, where the $F0$'s of the target words ''dog'' and ''bird'' were the same as the sentence in which they occurred. To produce the swapped condition, these target words were digitally swapped round at stop-closure silences between various combinations of files so that the target word did not have the same $F0$ as its carrier sentence. The swapped condition allows us to separate spatial and prosodic contributions to attention.

The resynthesised sentences and their targets were then given simulated reverberation using Peterson's ray-tracing model (Peterson, 1986) previously used by Culling *et al.* (1994). The model room layout is illustrated in Fig. 2 and is identical to that used by Culling *et al.* The source was simulated to be 2 m from the head, which was placed in a slightly different position in the room from that used by Culling *et al.* Positions of the source were chosen to give a direct path-length difference at the two ears corresponding to the ITDs of 0, ±45, ±91, and ±181 μs used in the earlier experiment. The model calculated the waveform at each of the two ears represented as points in free space. Consequently, the model does not represent either interaural intensity differences (IID) arising from head shadow or pinna effects. The model does, however, incorporate intensity differences arising from different path lengths from the source to the two ears, though for a source 2 m away and opposite to one ear the IID is small (c. 0.7 dB). For convenience, and for ease of comparison of the results with those from our previous experiments, we will refer to the different source positions by their corresponding ITDs. Two absorption coefficients, 0.6 and 0.3, were used to give reverberation times (RT_{60} is defined as the time for the reverberant energy to drop by 60 dB) of 0.14 and 0.4 s, respectively. Our directly measured reverberation times differ slightly from those reported in Culling *et al.* (1994).

B. Procedure

The 13 listeners were native speakers of British English aged between 21 and 52 (including the two authors); all had pure-tone thresholds within the normal range at octave frequencies between 125 Hz and 8 kHz. They had all participated in experiment 1 of Darwin and Hukin (1999).

The procedure was identical to that of experiment 1 of Darwin and Hukin (1999) except that listeners were told that they should attend to the sentence ''Could you please write the word X down now,'' and to press the ''d'' or ''b'' key if it contained the target word ''dog'' or ''bird,'' respectively. One carrier sentence and one target word always had an $F0$ of 100 Hz, the other carrier sentence and the other target had an $F0$ that was either the same or 1, 2, or 4 semitones higher. The attended carrier sentence was thus separated from the other sentence by seven different intervals (-4, -2, -1, 0, 1, 2, or 4 semitones).

For the trials on which the ITD was zero, these seven conditions were combined with two conditions in which the target word that had the same $F0$ as the attended sentence was either ''dog'' or ''bird'' giving a total of 14 conditions (two of which are in fact identical, with zero ITD and zero difference in $F0$).

For the trials on which the ITD was not zero, there were three values of ITD combined with: $F0$ difference (seven values), whether the target with the same ITD was ''dog'' or ''bird'' (two values), whether the attended sentence had a positive or a negative ITD (two values), whether the target word with the same ITD as the carrier sentence also had the same $F0$ as the carrier sentence or not (normal versus swapped: two values). This combination gives a total of 168 conditions (some identical) which were presented five times each with each listener getting a different pseudo-random order. All these trials were presented in separate blocks of trials at the two reverberation times (0.1 and 0.4 s) in a counter-balanced order across subjects. The sentences when mixed at each headphone (Sennheiser 414) gave an average level of 68 dB SPL through a flat-plate coupler.

C. Results and discussion

Listeners' preferences for one or the other target word were subjected to analysis of variance with the following factors: ITD (±45, ±91, ±181 μs), $F0$ difference between the attended carrier sentence and the distractor ($\Delta F0 = -4$, -2, -1, 0, $+1$, $+2$, $+4$ semitones), correct target (''dog,'' ''bird''), correct target's $F0$ relation to attended carrier (same, different), and side of attended sentence (left, right). The reported significance levels have had the Greenhouse–Geisser correction for sphericity applied using SuperANOVA (Abacus Concepts).

1. Continuity of F0

When the two carrier sentences and target words have the same, zero ITD, the only cue to which target word belongs with the attended carrier is $F0$. Figure 3 shows the percentage of target words reported that had the same $F0$ as the attended sentence for both the reverberation conditions of this experiment, and also for the same 13 listeners for experi-

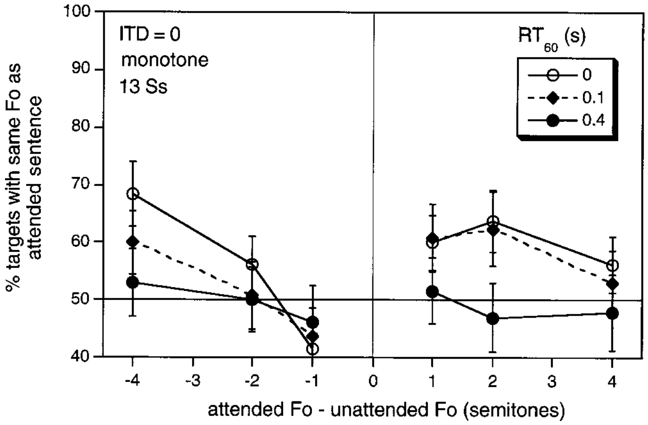

FIG. 3. Percent of target words reported that had the same $F0$ as the attended sentence for the conditions that had zero ITD in experiment 1 as a function of the $F0$ difference (attended-unattended) between sentences. The parameter is the reverberation time of the simulated room. Sound sources for both sentences were positioned directly ahead. Data for the $\mathrm{RT}_{60}=0$ condition are for the same subjects from experiment 1 of Darwin and Hukin (1999). There is no data point at 0 semitones since the correct response is undefined.

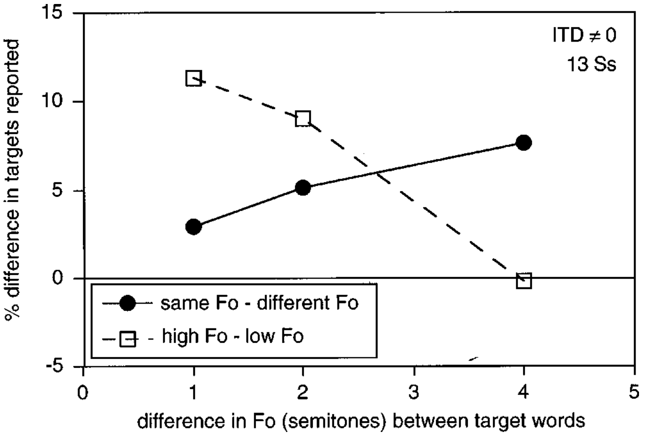

FIG. 4. The figure shows how strongly listeners in experiment 1 preferred the target word that had either the same (rather than different) $F0$ as the attended sentence (filled circles), or the target word that had the higher (rather than lower) $F0$ (open squares) as a function of the difference in $F0$ between target words (and so also between carrier sentences). Data are from trials where the ITDs were nonzero averaged across reverberation conditions.

ment 1 of Darwin and Hukin (1999) which used the same stimuli but without reverberation. The data are plotted as a function of the $F0$ difference between the attended and the carrier sentences.

Overall, the tendency for listeners to report the target with the same, monotonous $F0$ is rather weak, and is reduced by reverberation [$F(2,24)=15.8$, $p<0.0001$]. The reduction is rather slight for a RT_{60} of 0.1 s, but performance is at chance for a RT_{60} of 0.4 s. This result contrasts with Culling *et al.* (Exp. 3a, 1994), finding that a 0.4-s reverberation time did not reduce the substantial beneficial effect of a 1-semitone difference in $F0$ on the threshold level for identifying a steady-state vowel in the presence of a vowellike masker. Apart from the very different tasks used in the two experiments, possible factors responsible for the more substantial effect of reverberation time in the present experiment are the use of natural word targets embedded in a sentence context.

The below-chance performance for the two less reverberant conditions at an $F0$ difference of -1 semitone probably reflects a tendency for listeners to report the target word that has the higher $F0$ when there is only a small absolute difference in $F0$ between the two target words. This tendency is, in fact, a general one across the experiment. Figure 4 shows it for the rest of the experimental data (i.e., when there is also an ITD difference between the sentences). At small $F0$ differences, listeners tend to report the target word that has the higher $F0$, but at larger differences in $F0$ they tend to report the target word with the same $F0$ as the carrier sentence. Neither of these trends is large (maximally about 10%), but the interaction for the data in Fig. 4 is highly significant [$F(6,72)=40.4$, $p<0.0001$].

The tendency to prefer the higher of two different-pitched sounds has been noted previously. When two sentences are presented simultaneously with a pitch difference of 4 (but not at 1, 2, 6, or 8) semitones between them, the one with the higher $F0$ is more intelligible (Assmann, 1999a). Similarly, when listeners are asked to match the dominant pitch of two different vowels played simultaneously 2 or 4 semitones apart, listeners make matches which are closer to the higher than the lower pitch; when asked to match both pitches, matches tend to be made first and more accurately to the higher pitch (Assmann and Paschall, 1998). In a musical context, the higher of two polyphonic parts is easier to recognize than the lower (Gregory, 1990). However, extensive experiments on double-vowel recognition have failed to find any consistently better identification of the higher-pitched vowel (McKeown, 1992; de Cheveigné *et al.*, 1997; Paschall and Assmann, 1998). In our data, the preference for the higher-pitched target word weakens as the $F0$ difference increases from 1 to 4 semitones. This reduced preference for the higher-$F0$ target is probably due to an increased preference for the target with the same $F0$ as the attended sentence.

2. Continuity of ITD

When ITD is not zero, the spatial separation of the sentences provides an additional dimension for listeners to choose the target word. In the normal conditions spatial separation and $F0$ continuity work together whereas in the swapped conditions they are opposed. Since we have already shown that there is only a small tendency for listeners to report the target word with the same $F0$ as the attended carrier sentence, the data presented in Fig. 5 ignores the normal/swapped distinction. It shows the percentage of reported targets that had the same ITD as the attended sentence for the 13 Ss of experiment 1. The data in the left-hand panel of Fig. 5 are for the same subjects from experiment 1 of Darwin and Hukin (1999); they show that with no reverberation listeners show a strong tendency to report the target with the same ITD as the attended sentence.

The center and right-hand panels of Fig. 5 show that increasing reverberation time clearly reduces the ability of listeners to report the target that has the same ITD as the attended sentence [$F(2,24)=90.8$, $p<0.0001$]. The effect of

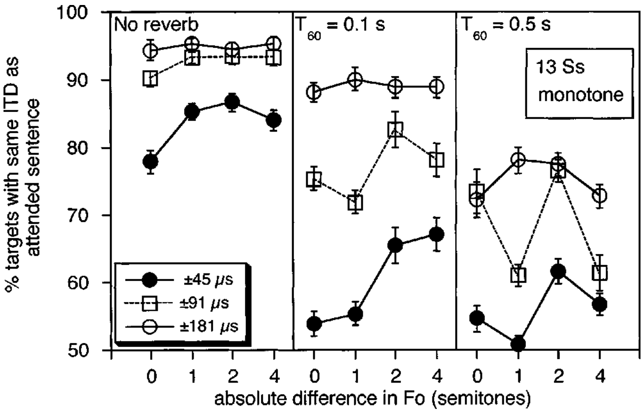

FIG. 5. Percent (±1 standard error) of target words reported that had the same ITD as the attended sentence for the 13 Ss of experiment 1. The $RT_{60}=0$ data are for these subjects from experiment 1 of Darwin and Hukin (1999).

reverberation is more marked for smaller ITDs [$F(4,48) = 22.5$, $p < 0.0001$], probably because of ceiling effects for the large ITD conditions. At the smallest ITD, ±45 μs, performance decreases markedly as reverberation time increases from 0 to 0.1 s, whereas for the largest ITD, listeners are still getting about 75% of the targets correct by ITD with a reverberation time of 0.4 s.

The data in Fig. 5 are shown plotted against the absolute difference in $F0$ between the two sentences; the generally small effects of the direction of $F0$ difference and of $F0$ continuity have been dealt with in the previous section. When there is no reverberation, subjects' ability to report the target with the same ITD as the attended sentence is somewhat lower when the two sentences have the same $F0$ than when there is a difference in $F0$. With increased reverberation, the effect of a difference in $F0$ becomes rather erratic, interacting with reverberation time and ITD [$F(24,288) = 2.3$, $p < 0.05$].

In summary, this experiment has shown that listeners' ability to use ITD to follow a particular target sentence is substantially disrupted when reverberation is introduced. In addition, the rather weak tendency to report the target that has the same $F0$ as the attended sentence, is also reduced by reverberation.

II. EXPERIMENT 2

The purpose of experiment 2 is to compare how resilient to reverberation are ITD and natural prosodic cues. The natural prosodic variation that we use here is more effective at maintaining a listener's attention than the monotone manipulations used in experiment 1 (Darwin and Hukin, 2000).

Reverberation impairs listeners' ability to use sinusoidally modulated $F0$'s to separate simultaneous vowellike sounds more than it impairs their ability to use monotonous $F0$'s (Culling *et al.*, 1994). Culling *et al.* attribute this impairment by reverberation to two causes: first, for small modulation depths reverberation blurs the harmonic structure of individual sounds; second, further impairment occurs when the modulation depth is comparable with the $F0$ separation of the two sounds so the blurred $F0$'s overlap. On the basis of this analysis we would expect reverberation to have more effect on the ability of listeners to use natural contours which are separated by an overall $F0$ difference which prevents the $F0$ of the reverberant sound becoming too close than on those with overlapping $F0$ contours.

A. Stimuli and procedure

The utterances used were those from experiment 1 of the companion article (Darwin and Hukin, 2000). In brief, two sentences with each of two target words (''Could you please write the word bead/globe down now'' and ''You'll also hear the sound bead/globe played here'') were spoken in two versions, one with the main sentence stress early in the sentence (on ''please'' or ''also''), and once with the stress late in the sentence (on ''now'' or ''here''). Sentences were paired so that a pair contained both carrier sentences, both target words, and both sentence stress positions. Three different resyntheses were then made for each sentence pair: *original*, in which the $F0$ values were unchanged; *together*, in which the two sentences, $F0$ contours were both shifted in order to make the values of $F0$ during the two target words similar; and *apart*, in which the two sentences, $F0$ contours were shifted the opposite way in order to make their values of $F0$ during the two target words more different. The $F0$ contours for a representative pair of sentences are shown in Fig. 1 of Darwin and Hukin (2000).

In order to measure the relative strengths of the ITD and prosodic cues, two different conditions were generated from these resynthesized sentence pairs: a normal condition, which retained the sentences as described in the previous paragraph, and a swapped condition in which the target words (with their prosodic attributes) were swapped between the two sentences of a pair. This swapping of target words was done before sentences were given different ITDs. A target word that had the same ITD as the target sentence would thus also have the same prosody in the normal condition, but different prosody in the swapped condition.

The simulated room used in experiment 1 with the same spatial positions and with $RT_{60}=0.4$ s then produced stereo sound files for presentation to listeners using the same procedure as in the previous experiment.

The experiment was taken by the 13 listeners who had taken the equivalent nonreverberant experiment—experiment 1 in the companion article (Darwin and Hukin, 2000). Half of the listeners were instructed to listen to one sentence throughout the experiment, and the others were instructed to listen to the other sentence.

B. Results and discussion

The percentage of targets reported that have the appropriate prosody for the attended sentence are shown in Fig. 6.

For the normal conditions, targets that have the appropriate prosody also have the same ITD as the attended sentence. Since both cues favor the same target word, performance is very high (about 90%) despite the reverberation both for the original condition where the $F0$ contours are unchanged, and for the apart condition where the $F0$ contours have been further separated by about 4 semitones. Performance goes down to about 70% in the together condition, where the two target words have very similar $F0$ contours.

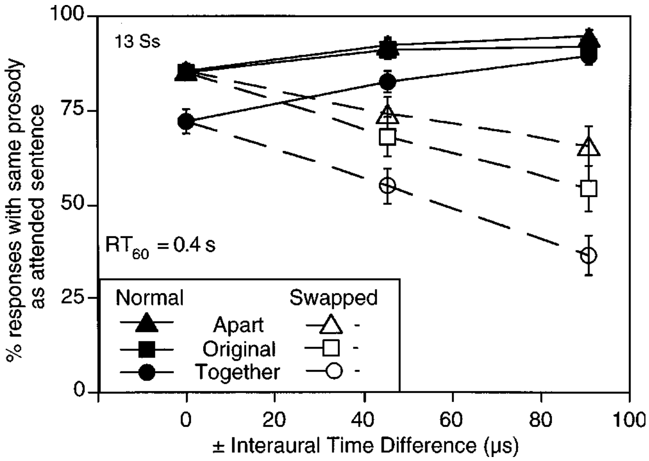

FIG. 6. Percent (±1 standard error) of target words reported that had the same prosody as the attended sentence as a function of the difference in ITD between the two sentences for the 13 Ss of experiment 2. Symbols joined by solid lines plot data from conditions where the target word with the same prosody as the attended sentence also had the same ITD. For those joined by dashed lines, the target word with the same prosody as the attended sentence had the same ITD as the unattended sentence.

Listeners are here mainly relying on other prosodic differences between the two words such as the level differences described in the companion article.

For the swapped conditions, the spatial cues are placed in opposition to the prosodic cues. As the nominal ITD increases, listeners report more of the targets that have the appropriate spatial cues and correspondingly fewer targets that have the correct prosody. In order to see more clearly the relative effect of reverberation on the two types of cue, Fig. 7 shows the size of the decrease, between the normal and the swapped conditions, in the percentage of reported targets that have the same ITD as the attended sentence. The more effective the prosodic cue, or the less effective ITD, the larger will be this change.

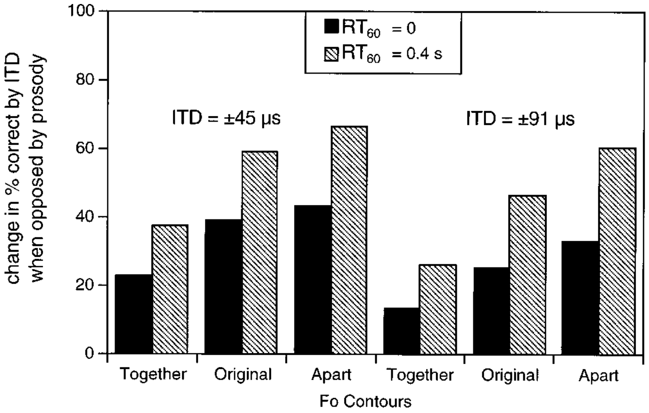

FIG. 7. Influence of reverberation on the relative effectiveness of ITD and prosody in experiment 2. The bars indicate the decrease in the percent of target words that had the same ITD as the attended sentence between the normal conditions (in which the target word shared both ITD and prosody with the attended sentence) and swapped conditions (where the target word had the same ITD but a different prosody). A large score indicates that prosodic cues are dominating over ITD. The scores are increased by increasing the $F0$ difference between the sentences (abscissa) and by reducing ITD. The increase with reverberation (shaded versus black bars) indicates that reverberation is degrading the ITD cue more than it is degrading the prosodic cue.

The solid bars of Fig. 7 show data from experiment 1 in the companion article, and their pattern simply shows that the *relative* effectiveness of the prosodic cues is increased by separating the $F0$ contours, and decreased by the larger ITD, as one would expect. The effect of reverberation (hatched bars) on this pattern is generally to favor the prosodic cues indicating that 0.4-s reverberation weakens the spatial cue more than it does the prosodic ones. That the $F0$ component of the prosody maintains its effectiveness compared with the spatial cue is shown by the fact that the hatched bars in Fig. 7 remain above the solid bars in the original and apart $F0$ conditions.

III. EXPERIMENT 3

The final experiment examines the effect of reverberation on listeners' ability to attend to a sentence which is spoken by a different apparent vocal-tract size from the unattended sentence. The experiment takes the original prosody conditions from experiment 2 and introduces a ±15% difference in spectral envelope between the attended and unattended sentences and also between the two target words.

A. Stimuli and procedure

The original sentences from experiment 2 were modified to produce two apparently different talkers by altering the spectral envelope without changing the $F0$ or duration. Sentences were resynthesised using PSOLA using a method described in detail in Darwin and Hukin (2000). Briefly sentences were globally reduced in duration and raised in $F0$ by 15%, resampled at 15% higher sampling frequency, and then played back at the original sample rate. This manipulation produced sentences that had the same durations and $F0$'s as the originals, but which had the spectral envelope lowered by 15% (effectively increasing the apparent vocal-tract size). Shorter vocal-tract voiced sentences were produced using opposite manipulations. A large difference in vocal-tract length was used since earlier work (Darwin and Hukin, 2000) had shown that such large differences were necessary to influence attention substantially. The vocal-tract was both lengthened and shortened by 15% (rather than being merely shortened by 30%) in order to maintain speech quality.

The sentences were paired as in experiment 2 with the additional constraint that each pair contained one long vocal-tract sentence and one short vocal-tract sentence. Target words could be swapped between the sentences of a pair before the sentences were allocated an ITD. In the swapped condition, the target word had the same ITD as the attended sentence, but the vocal-tract size and prosody of the unattended sentence.

Reverberation was added to the sentences in each pair using the same simulated room used in experiments 1 and 2, the same spatial positioning and with $\mathrm{RT}_{60} = 0.4\,\mathrm{s}$. The resultant stereo sound files were presented to the listeners using an identical procedure to the previous two experiments. Half of the listeners were instructed to listen to one sentence throughout the experiment, and the others were instructed to listen to the other sentence.

Of the 11 listeners who took the experiment, 8 had already taken part in the corresponding experiment in the com-

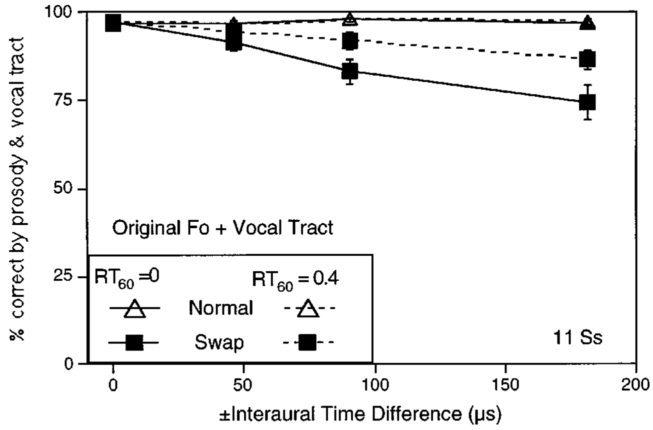

FIG. 8. Percent (±1 standard error) of target words reported that had the same prosody and vocal tract size as the attended sentence as a function of the difference in ITD between the two sentences for the 11 Ss of experiment 3. The open triangles plot data from conditions where the target word with the same prosody and vocal-tract size as the attended sentence also had the same ITD. The filled squares plot data from conditions where the target word with the same prosody and vocal-tract size as the attended sentence had the same ITD as the unattended sentence.

panion article with no reverberation. The remaining 3 subjects additionally took this experiment so that we had data from all 11 subjects both with and without reverberation.

B. Results and discussion

Introducing a difference in vocal-tract size between the attended and unattended sentences provides a cue which is extremely resistant to degradation by reverberation even when opposed by a substantial difference in ITD. Figure 8 shows the percent of target words reported that had the same prosody and vocal tract size as the attended sentence as a function of the difference in ITD between the two sentences. The notable feature of the figure is that in the swapped conditions, where the target that is correct by prosody and vocal-tract size has the wrong ITD, listeners very strongly prefer the target that has the correct prosody and vocal-tract size. This preference is still very substantial (87%) even when an ITD difference of ±181 μs opposes these cues.

The substantial preference for the target word that had the same vocal-tract size as the carrier sentence contrasts with recent findings by Assmann (1999b) who found little consistent improvement in overall word intelligibility of simultaneously presented sentences with differences in vocal-tract size. Two differences between our experiments are probably responsible. First, we used a difference in vocal-tract length (−15% vs +15%) that is almost twice that used by Assmann (0% vs ±20%); smaller differences than our ±15% are less effective as cues (Darwin and Hukin, 2000). Second, as pointed out in the Introduction, while Assmann's experimental paradigm measures overall word intelligibility, but does not measure attention, our paradigm measures attention but does not measure word intelligibility.

IV. SUMMARY AND GENERAL DISCUSSION

The three experiments reported here have explored how resistant to simulated reverberation are various cues for selective attention. The listeners' task was to decide which of two simultaneous target words belonged to an attended rather than an unattended sentence. The experiments have shown the following.

(i) The weak preference of listeners for the target word with the same monotonous $F0$ as the attended sentence was eliminated by reverberation (experiment 1).

(ii) Listeners' ability to use ITD differences was seriously impaired by reverberation (experiment 1), although some ability remained for the longest ITD tested.

(iii) The prosodic manipulations (including $F0$ differences) used in experiment 2 were generally more resistant to reverberation than were the ITD differences.

(iv) Adding a difference in vocal-tract size to the prosodic cues produced a high level of performance which was very resistant to reverberation (experiment 3).

This paper has explored the effect of reverberation on one aspect of auditory attention: the ability of listeners to maintain attention to a particular sound source across time. We have used a paradigm that has minimal intelligibility requirements since listeners are always presented with the same two target words. The paradigm also minimizes local cues to sound-source continuity by presenting the two target words simultaneously and with silence (from stop closure) before and after. Although it might be argued that this latter constraint is somewhat artificial, it perhaps reflects the practicalities of attempting to assign to the appropriate sources, speech which is intermittently masked by other talkers.

The two main results of the paper are, first, that reverberation reduces the effectiveness of ITD as an attentional cue, and, second, that natural prosodic cues and large vocal-tract size differences between talkers provide additional cues that are generally more resistant to the effects of reverberation than is ITD.

Our previous work had shown that, in the absence of talker and $F0$ differences, listeners could use small differences in ITD to follow a particular sound source. The effectiveness of this cue (with monotone speech), however, is seriously reduced by reverberation despite the dynamic nature of the sentences and target sounds. Reverberation alters both the intensity and the phase of individual harmonics and also reduces their depth of amplitude modulation and so will make both the carrier sentence and the target harder to localize. Since the detrimental effect of reverberation persists even when there is a 4-semitone difference in $F0$ between sentences, it is unlikely that interference between the individual components of the two sentences is necessary for reverberation to have its effect.

Natural prosodic cues, although not particularly strong (at least in these experiments) in the absence of reverberation, are more resistant to reverberation than are the spatial cues used here. Both $F0$ differences and other prosodic cues (probably mainly amplitude differences) help listeners to select the appropriate target word, and their influence becomes relatively stronger than the spatial cue with increased reverberation. A large difference in vocal-tract size provides a more powerful continuity cue, and this too is very resistant to reverberation.

Although the paradigm used here allows sophisticated stimulus manipulations to be made easily with natural stimuli, it does not allow us to determine to what extent the cues we are investigating are having their effect in real time, and to what extent they are (merely) influencing a listener's choice some time after the event. It is possible that spatial cues could be exploited more easily than prosody or vocal-tract size for directing of attention (Spence and Driver, 1994). If this were the case, then a task such as shadowing that is more sensitive to the real-time allocation of attention might be substantially disrupted by the erosion of spatial cues by reverberation even though prosodic and vocal-tract cues were also present.

ACKNOWLEDGMENTS

This research was supported by Grant No. G9801285 from the UK MRC to the first author. We are grateful to John Culling for help with the software implementation of the reverberation. Peter Assmann and Alain de Cheveigné made a number of helpful suggestions for improving the paper.

Abouchacra, K. S., Emanuel, D. C., Blood, I. M., and Letowski, T. R. (**1998**). ''Spatial perception of speech in various signal to noise ratios,'' Ear Hear. **19**, 298–309.

Assmann, P. F. (**1999a**). ''Fundamental frequency and the intelligibility of competing voices,'' 14th International Congress of Phonetic Sciences, San Francisco, 1–7 August 1999, pp. 179–182.

Assmann, P. F. (**1999b**). ''Vocal tract size and the perception of competing voices,'' J. Acoust. Soc. Am. **106**, 2272.

Assmann, P. F., and Paschall, D. D. (**1998**). ''Pitches of concurrent vowels,'' J. Acoust. Soc. Am. **103**, 1150–1160.

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

Bird, J., and Darwin, C. J. (**1998**). ''Effects of a difference in fundamental frequency in separating two sentences,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Culling, J. F., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels: Within and across-formant grouping by $F0$,'' J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1994**). ''Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels,'' Speech Commun. **14**, 71–95.

Darwin, C. J., and Hukin, R. W. (**1999**). ''Auditory objects of attention: the role of interaural time-differences,'' J. Exp. Psychol. **25**, 617–629.

Darwin, C. J., and Hukin, R. W. (**2000**). ''Effectiveness of spatial cues, prosody and talker characteristics in selective attention,'' J. Acoust. Soc. Am. **107**, 970–977.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997**). ''Concurrent vowel identification. 1. Effects of relative amplitude and F-0 difference,'' J. Acoust. Soc. Am. **101**, 2839–2847.

Good, M., and Gilkey, R. (**1996**). ''Sound localization in noise: The effect of signal-to-noise ratio,'' J. Acoust. Soc. Am. **99**, 1108–1117.

Gregory, A. H. (**1990**). ''Listening to polyphonic music,'' Psychol. Music **18**, 163–170.

Hartmann, W. M. (**1983**). ''Localization of sound in rooms,'' J. Acoust. Soc. Am. **74**, 1380–1391.

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.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). ''Sound localization in noise in normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 1810–1820.

McKeown, J. D. (**1992**). ''Perception of concurrent vowels: the effect of varying their relative level,'' Speech Commun. **11**, 1–13.

Moncur, J., and Dirks, D. (**1967**). ''Binaural and monaural speech intelligibility in reverberation,'' J. Speech Hear. Res. **10**, 186–195.

Moulines, E., and Charpentier, F. (**1990**). ''Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones,'' Speech Commun. **9**, 453–467.

Nabelek, A. K. (**1988**). ''Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss,'' J. Acoust. Soc. Am. **84**, 476–784.

Nabelek, A. K., and Dagenais, P. A. (**1986**). ''Vowel errors in noise and in reverberation by hearing-impaired listeners,'' J. Acoust. Soc. Am. **80**, 741–748.

Nabelek, A. K., and Donahue, A. M. (**1984**). ''Perception of consonants in reverberation by native and non-native listeners [letter],'' J. Acoust. Soc. Am. **75**, 632–634.

Nabelek, A. K., and Robinson, P. K. (**1982**). ''Monaural and binaural speech perception in reverberation for listeners of various ages,'' J. Acoust. Soc. Am. **71**, 1242–1248.

Paschall, D. D., and Assmann, P. F. (**1998**). ''Ranking the pitches of concurrent vowels,'' J. Acoust. Soc. Am. **103**, 2980.

Peterson, P. M. (**1986**). ''Simulating the response of multiple microphones to a single source in a reverberant room,'' J. Acoust. Soc. Am. **80**, 1527–1529.

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.

Plomp, R. (**1977**). ''Acoustical aspects of cocktail parties,'' Acustica **38**, 186–191.

Rakerd, B., and Hartmann, W. M. (**1985**). ''Localization of sound in rooms, II: The effects of a single reflecting surface,'' J. Acoust. Soc. Am. **78**, 524–533.

Scheffers, M. T. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Ph.D. dissertation, Groningen University, The Netherlands.

Spence, C. J., and Driver, J. (**1994**). ''Covert spatial orienting in audition: exogenous and endogenous mechanisms,'' J. Exp. Psychol. **20**, 555–574.