Effects of reverberation on perceptual segregation of competing voices

John F. Culling,[a] Kathryn I. Hodder, and Chaz Yee Toh
School of Psychology, Cardiff University, P.O. Box 901, Cardiff CF10 3YG, United Kingdom

(Received 19 September 2002; revised 5 August 2002; accepted 6 August 2003)

Two experiments investigated the effect of reverberation on listeners' ability to perceptually segregate two competing voices. Culling *et al.* [Speech Commun. **14**, 71–96 (1994)] found that for competing synthetic vowels, masked identification thresholds were increased by reverberation only when combined with modulation of fundamental frequency ($F0$). The present investigation extended this finding to running speech. Speech reception thresholds (SRTs) were measured for a male voice against a single interfering female voice within a virtual room with controlled reverberation. The two voices were either (1) co-located in virtual space at 0° azimuth or (2) separately located at ±60° azimuth. In experiment 1, target and interfering voices were either normally intonated or resynthesized with a fixed $F0$. In anechoic conditions, SRTs were lower for normally intonated and for spatially separated sources, while, in reverberant conditions, the SRTs were all the same. In experiment 2, additional conditions employed inverted $F0$ contours. Inverted $F0$ contours yielded higher SRTs in all conditions, regardless of reverberation. The results suggest that reverberation can seriously impair listeners' ability to exploit differences in $F0$ and spatial location between competing voices. The levels of reverberation employed had no effect on speech intelligibility in quiet. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1616922]

PACS numbers: 43.66.Pn, 43.66.Dc, 43.55.Hy [LRB] Pages: 2871–2876

I. INTRODUCTION

Most research on the perceptual effects of reverberation on speech has concentrated upon its effects on the transmission of a single voice in quiet (Houtgast and Steeneken, 1985). This work has been applied, in the form of the speech transmission index, to the particular problems of theatrical auditoria and lecture rooms where one-way verbal communication is the norm. Often these spaces are large and a degree of reverberation is desirable as a means of delivering the necessary sound level to the audience. However, too much reverberation can smear the temporal envelope of the speech, ultimately rendering it unintelligible. The speech transmission index can be used to predict the intelligibility of speech in quiet (or in simple forms of noise, such as might be produced by air conditioning) in different environments.

A relatively small amount of research has been conducted on the effects of reverberation on multi-talker communication (Plomp, 1976; Culling *et al.*, 1994; Darwin and Hukin, 2000). However, such work as exists seems to have serious implications for room design, because reverberation disrupts listeners' ability to cope with multiple overlapping voices far more easily than it does the intelligibility of a voice in quiet. Plomp used a reverberation room with varying amounts of inserted sound-absorbing material to show that thresholds for speech reception against interfering speech or noise were increased in a more reverberant enclosure. Furthermore, the beneficial effect of spatial separation of the target and interfering sources was largely abolished in the presence of reverberation. Culling *et al.* measured the masked identification thresholds for synthesized vowel sounds in a virtual-acoustic space with controlled surface absorption. Using a pink noise masker, they also found that the effect of spatial separation was easily abolished by reverberation. Using a competing vowel as a masker, the beneficial effect of differences in fundamental frequency ($F0$) between the two vowels (Scheffers, 1983; Assmann and Summerfield, 1990; Culling and Darwin, 1993a) was robust to reverberation. However, sinusoidal modulation of $F0$ (for both target and masker), which in anechoic conditions had no effect on masked thresholds, resulted in the abolition of the $F0$-difference benefit when combined with reverberation. Darwin and Hukin used a similar virtual-acoustic method to examine the effects of reverberation on listeners' ability to track a particular voice over time. They found that reverberation substantially reduced listeners' ability to use interaural time delays to attribute competing words to the correct carrier sentences. However, for this task, the benefits of continuity of $F0$ and vocal tract length were more robust in reverberation.

Current theories of segregation by $F0$ suggest that the auditory system can suppress one harmonic interfering voice, perhaps by a harmonic-cancellation process (de Cheveigné, 1997), permitting superior understanding of the remaining voice. The evidence for this scheme is largely based on experiments with simultaneous vowels. If simultaneous vowels have different $F0$s, then they can be identified more accurately than if they have the same $F0$, but two different $F0$s are not the only form of excitation of the vowels that will result in improvements in their identification. It is evidence from these alternative forms of excitation that points specifically to cancellation. If one vowel is inharmonic (Summerfield and Culling, 1992; de Cheveigné *et al.*, 1997), recognition of the competing vowel is improved compared to having

[a]Electronic mail: cullingj@cardiff.ac.uk

both on the same $F0$. Similarly, if one vowel is excited by noise, resulting in a whispered timbre, recognition of that vowel improves compared to having both vowels on the same $F0$ (Lea, 1992). In both cases, identification of the vowel that remains harmonic is unaffected. However, if both vowels are inharmonic or whispered, the advantage is lost. Thus, if one vowel has any spectral microstructure that differs from a harmonic competitor, then it can be perceptually separated from that competitor and better identified. A cancellation mechanism would be expected to produce this pattern of performance, because it will cancel the harmonic competitor, leaving inharmonic or noise-excited sounds relatively unaffected. In real listening situations, both voices are harmonic, but (most of the time) differ in $F0$. It seems likely that the cancelled voice is the dominant and/or more intense one, because ability to match the pitch of each vowel correlates with identification accuracy (Assmann and Paschall, 1998) and identification of the $F0$ is presumably a prerequisite for cancellation.

The human voice varies rapidly in $F0$ over a full octave during normally intonated speech. The question therefore arises of how the cancellation mechanism deals with this moving target. Further experiments with simultaneous vowels have modulated $F0$ sinusoidally, creating an effect similar to operatic vibrato. Using these stimuli, it has been found that the ability to exploit differences in $F0$ seems to correlate with the mean instantaneous difference in $F0$ across the stimulus (as opposed to the difference between the long-term mean $F0$'s). Thus, vowels modulated out of phase around the same mean $F0$ are better identified than if they are modulated in the same phase (Darwin and Culling, 1990).

Harmonic cancellation of the dominant voice will provide the listener with better identification of individual speech sounds, but the reconstruction of separated sentences also requires the linkage of separated speech elements across time. In addition to this cancellationlike process, therefore, it is possible that listeners use $F0$ in a number of other ways. First, the mean $F0$ of a person's voice may be used in order to focus attention on that voice in the presence of a competing interferer with a different mean $F0$ (Cherry, 1953); this would enable a listener to acquire or reacquire the appropriate stream of information and to avoid confusing it with the interfering stream. Second, the attention on the correct stream can also be maintained if the $F0$ of the target voice is tracked continuously (Parsons, 1976). Continuous tracking of the $F0$ may enable a listener to deal with two voices with the same mean $F0$, although the tracking process is susceptible to confusion when the two voices' $F0$'s intersect (Culling and Darwin, 1993b). Darwin and Hukin's (2000) experiments with reverberation indicate that use of the $F0$ contour to track a target voice is also affected by reverberation, but that it is more robust to reverberation than benefits due to differences in spatial location.

It is not entirely clear how the combination of $F0$ modulation and reverberation disrupts these $F0$-segregation mechanisms. However, it seems likely that, when the $F0$ varies over time, wavefronts that have been delayed by their passage around the walls of the room have a different $F0$ from direct sound that arrives simultaneously at the receiver; direct and reflected waves were emitted from the modulating source at different times. In this way, the $F0$ of the interfering source is smeared in the sense that the harmonic series is less clearly defined in the stimulus. This smearing may make the interfering voice more difficult to cancel. Darwin and Hukin (2000) showed that reverberation can also upset listeners ability to use $F0$ in order to link successive words from the same voice. It is less clear how the smearing would affect that process.

The present investigation is a follow up to that of Culling *et al.* (1994), using running speech. The stimuli used in their study were highly artificial, but contained key features found in everyday listening situations. Some degree of reverberation is common to practically all listening environments and modulation of $F0$, while not normally sinusoidal, is unavoidable in natural speech. Indeed, normally intonated speech involves modulation of $F0$ that is both rapid (up to 5 oct/s) and typically varies over a full octave (O'Shaunessy and Allen, 1983). This modulation is both faster and more extreme than the $\pm 0.7\%-12\%$, 5-Hz sinusoidal modulation used by Culling *et al.* It is noteworthy that the combination of such subtle modulation of $F0$ and reverberation resulted in a collapse in listeners' ability to use differences in $F0$, one of the best-established cues to perceptual separation of competing voices.

II. EXPERIMENT 1

Because Culling *et al.* (1994) found that the effect of differences in $F0$ was robust to reverberation when $F0$ was not modulated, but not when it was modulated, experiment 1 tested whether the same happens with running speech. In order to do this, the speech was resynthesized with either the original or a monotonized $F0$ contour. This method has previously been used in order to control differences in $F0$ for concurrent speech (Brokx and Nooteboom, 1982). Then, speech reception thresholds (SRTs) were measured for target and interfering voices that had these different contours. The results of Culling *et al.* suggest that SRTs might be lower using the monotonized speech than using normally intonated speech when reverberation is present, because the $F0$-segregation mechanism is only impaired in the intonated case.

A. Stimuli

The corpus of sentences was from the Harvard Sentence List (Rothauser *et al.*, 1969). The recordings of voice DA, made at M.I.T. and digitized at 20 kHz with 16-bit quantization, were used as the basis of all stimuli. The sentences have low predictability and each has five designated keywords (given here in capitals). For instance, one sentence used in the current experiment was ''the STEMS of the TALL GLASSES CRACKED and BROKE.'' These sentences were manipulated using the Praat PSOLA speech analysis and resynthesis package. For monotonized speech, the mean $F0$ for each sentence was calculated and the sentence was resynthesized with this $F0$ throughout.

Interfering sentences were generated by feminizing the voice of DA. His voice was increased in $F0$ by a factor of 1.8 and, using the resynthesis and resampling[1] method of

Darwin and Hukin (2000), the spectral envelope was shifted up in frequency by 15%, to simulate a shorter vocal tract. The factor of 1.8 reduced the number of target/interferer pairs for which the monotonous versions were an exact octave apart; the resulting mean absolute deviation from an octave relationship was just over 2 semitones, but 15% were still within half a semitone of an octave difference. Eight interfering sentences were created in this way.

Reverberation was added using the image (ray-tracing) method (Allen and Berkley, 1979; Peterson, 1986) as implemented in the |WAVE signal processing package (Culling, 1996). The virtual room and source/receiver configuration was identical to that of Culling *et al.* (1994). The room had dimensions 5 m long×3.2 m wide×2.5 m high and virtual sources were 2 m from the receivers. The two receivers, separated by 20 cm, were placed along an axis at 30° to the 5-m wall on either side of a center point located 1.2 m from the 5-m wall and 1.9 m from the 3.2-m wall. The receivers were modeled as microphones suspended in space with no head between them. Absorption coefficients for each internal surface of the room were 0.3 for the reverberant room, giving a direct-to-reverberant ratio of −10 dB and a reverberation time of approximately 400 ms. For the anechoic room the coefficients were set to 1, giving an infinite direct-to-reverberant ratio. Binaural stimuli were produced by generating the impulse responses for the two receivers in virtual space and convolving the speech samples with these two impulse responses.

Stimuli were created for eight different conditions. These conditions covered two levels of reverberation (anechoic versus reverberant), two forms of intonation (original versus monotonized) and two spatial configurations (0°/0° vs +60°/−60°) in all possible combinations. Ten target sentences were created for each condition. Target and interfering sounds shared the same reverberation and form of intonation.

B. Procedure

Sixteen listeners each attended a single 90-min session. The session began with two practice runs using monaurally presented and unprocessed speech, in order to familiarize the listeners with the task. The following eight runs measured SRTs in each of the eight different conditions. The order of the conditions was rotated for successive listeners, while the sentence materials remained in the same order. Each of the 80 target sentences was thus presented to every listener in the same order and contributed equally to each condition. This procedure also ensured that each condition was presented in each serial position within the experimental session, counterbalancing order effects.

SRTs were measured using a 1-up/1-down adaptive threshold method (Plomp and Mimpen, 1979; Plomp, 1986; Culling and Colburn, 2000). For an individual SRT measurement, the ten male-voice target sentences were presented one after another, each one against the same ''female-voice'' interfering sentence. The listeners were instructed to listen to the male voice. The target-to-interferer ratio was initially very low. In the initial phase, listeners had the opportunity to listen to the first sentence a number of times, each time with an increased target-to-interferer ratio. Listeners were instructed to attempt a transcript of the first sentence using a computer terminal when they believed that they could hear more than half the words of the male voice. Once the first transcript was entered, the correct transcript was displayed on the computer terminal, with the five key words in capitals. The listener self-marked how many of the key words were correct. Subsequent target sentences were presented only once and self-marked in a similar manner; the level of the target speech was decreased by 2 dB if the listener correctly identified three or more of the five key words in the previous sentence, and otherwise increased by 2 dB. SRTs for a given condition/run were taken as the mean signal level derived in this way on the last eight trials. Each SRT measurement used a different interfering sentence.

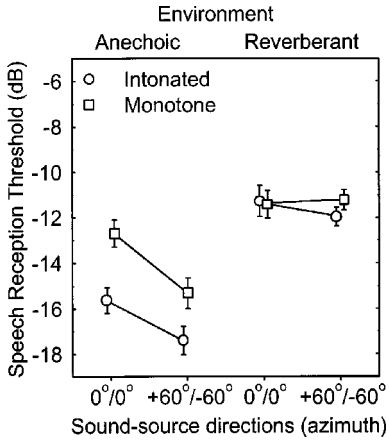

FIG. 1. Mean speech-reception thresholds in anechoic and reverberant conditions and for intonated (circles) and monotonous (squares) speech. Target and interfering sources were either both in front (0°/0°) or on either side (+60°/−60°). Lower thresholds imply greater intelligibility and/or perceptual separation. Error bars are one standard error of the mean.

Signals were digitally mixed, D/A converted, and amplified using a Tucker-Davis System II psychoacoustic rig (AP2, DD1, PA4, HB6) and presented to listeners over Sennheisser HD414 headphones in a single-walled IAC sound-attenuating booth within a sound-treated room. A computer terminal screen was visible outside the booth window; its keyboard was inside.

C. Results

In the anechoic conditions, Fig. 1 shows that mean SRTs were lower for intonated speech, indicating that listeners found the intonated speech intrinsically more intelligible than the monotonized speech. However, in the reverberant conditions there was no such effect. A three-factor analysis of variance (environment×$F0$ contour×spatial separation) reflected this pattern with a significant main effect of $F0$ contour $[F(1,15)=10.4, p<0.01]$ and an interaction between environment and $F0$ contour $[F(1,15)=20.0, p<0.001]$. Similarly, SRTs were lower for spatially separated voices in anechoic conditions, but not in reverberant conditions, producing a significant main effect of spatial separation $[F(1,15)=14.7, p<0.005]$ and an interaction between spatial separation and environment $[F(1,15)=5.4, p<0.05]$. Thus listeners could exploit the differences in spa-

tial location between the two competing voices, but only in the anechoic case. As a result of both these environment-specific effects, SRTs were also significantly lower overall in anechoic conditions [$F(1,15)=112.2$, $p<0.001$].

D. Discussion

Consistent with the results of Plomp (1976) and Culling *et al.* (1994), reverberation abolished listeners' ability to exploit differences in spatial location. The effect was more complete in the present study (and in that of Culling *et al.*) than in Plomp's experiments, probably because the sound sources in the present study were placed at a greater distance (2 m vs 1 m), within a smaller (virtual) room of (40 m^3 vs 63 m^3); both of these factors would have the effect of reducing the direct to reverberant ratio. This result is also consistent with Hukin and Darwin's work (using a similar virtual room, but with slightly different listener position) on the roles of $F0$ and ITD regarding the specific task of linking words from the same utterance. They found that reverberation disrupted both cues, but that the usefulness of different ITDs was more easily disrupted by reverberation than the effects of different $F0$s.

However, contrary to expectations based on the results of Culling *et al.*, monotonous speech was no more intelligible than intonated speech under reverberant conditions. In fact, intonated speech gave lower SRTs than monotonous speech in anechoic conditions and the two were approximately equal in reverberant conditions. It is possible that the monotonous condition was impaired to some degree by the occasional pair of target and interfering sentences that were close to an octave relationship. However, the advantage of intonated speech in anechoic conditions can probably be related more to exploitation of prosodic information. Prosodic information is provided by variations in the $F0$, amplitude, and rhythm of speech, so monotonization removes one of these three sources of information. The information contributes to intelligibility at multiple levels (Cutler *et al.*, 1997) and the removal of the $F0$-modulation element produces a cost in intelligibility equal to a 2.5-dB change in SRT (based on the difference in SRT for anechoic monotone and intonated $F0$ contours).

Since intonated speech is intrinsically more intelligible in anechoic conditions, one interpretation of the pattern of results is that reverberation destroys listeners' ability to exploit prosodic information conveyed by the intonation contour to assist speech intelligibility. However, given the results from Culling *et al.*'s experiments with concurrent synthetic vowel sounds, there is a more likely interpretation. It may be that intonated speech is intrinsically more intelligible than monotonous speech for all conditions, but that it is difficult to use $F0$ differences to perceptually separate two intonated voices in a reverberant setting; the monotonous speech may be perceptually separated from the (monotonous) interfering voice relatively well in the reverberation, but since it is less intelligible than the intonated speech, the SRT is no better. These two effects may be offsetting each other and yielding similar SRTs in all the reverberant conditions. Experiment 2 was designed to differentiate between these two possibilities.

III. EXPERIMENT 2

Experiment 2 discriminated between the different possible interpretations of the results from experiment 1 by adding conditions that used inverted intonation contours. These contours provide equal modulation of $F0$ (to disrupt segregation by $F0$ under reverberation), but were not expected to contribute to intrinsic speech intelligibility. Speech with an inverted $F0$ contour has a vague, questioning tone; the fall in $F0$ characteristic of the end of a statement is replaced with the rising $F0$ contour of a question and the stress sounds odd, because stressed syllables have an unnatural combination of low pitch and high intensity. Otherwise, the inverted-$F0$ speech sounded clearly articulated and natural.

A. Stimuli

The stimuli were largely similar to those for experiment 1, but using different target sentences from the same voice. In addition, the larger number of conditions required some additional interfering sentences; the choice of all 12 interfering sentences was reviewed to ensure that they were longer than all target sentences.

The eight conditions from experiment 1 were replicated. Four additional conditions were added that had inverted $F0$ contours. Inversion of the $F0$ contour was applied to both target and interferer. For inverted speech, the new $F0$, $F0'$, was derived for each analysis frame using the following equation:

$$F0' = \frac{\text{mean}\, F0^2}{F0}. \tag{1}$$

Here, $F0$ is the fundamental frequency of the frame and mean $F0$ is the mean fundamental frequency calculated over the duration of the sentence.

B. Procedure

Thirty-six new listeners each attended a single 2-h session. They completed the same two practice runs as in experiment 1 and 12 experimental runs, covering the 12 different conditions. As in experiment 1 the conditions were rotated from one listener to the next, while the sentence materials remained in the same order. The equipment was identical save for the use of Sennheiser HD590 headphones.

C. Results

Figure 2 shows mean SRTs for 36 listeners in experiment 2. SRTs for the eight conditions replicated from experiment 1 were similar in pattern to those from that experiment, although on average several dB higher. The effect of spatial location was, again, abolished by reverberation, and intonated speech again gave lower thresholds than monotonous speech in anechoic conditions only. SRTs for the four additional conditions with inverted $F0$ contours were substantially higher than the other conditions across all conditions of reverberation and spatial separation.

The results were analyzed with a three-way analysis of variance (environment$\times F0$ contour$\times$spatial separation). SRTs were, again, significantly lower for spatially separated sources [$F(1,35)=26.9$, $p<0.001$] and under anechoic con-

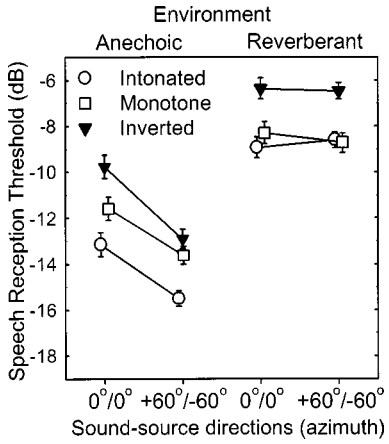

FIG. 2. Mean speech-reception thresholds in anechoic and reverberant conditions and for intonated (open circles), monotonous (open squares), and inverted-$F0$ (filled inverted triangles) speech. Target and interfering sources were either both in front (0°/0°) or on either side (+60°/−60°). Lower thresholds imply greater intelligibility and/or perceptual separation. Error bars are one standard error of the mean.

ditions [$F(1,35)=407.5$, $p<0.001$]. In addition, the three types of $F0$ contour differed significantly [$F(2,70)=60.3$, $p<0.001$]. The effects of spatial separation and $F0$ contour interacted with the presence of reverberation. First, the effect of spatial separation was abolished in reverberation, producing an interaction between environment and spatial separation [$F(1,35)=51.1$, $p<0.001$]. Simple main effects showed the effect of spatial separation to be significant only in anechoic conditions [$F(1,1)=50.9$, $p<0.001$]. Second, the convergence of SRTs from the monotonous and normally intonated conditions under reverberation produced an interaction between $F0$ contour and environment [$F(2,70)=4.3$, $p<0.02$]. However, SRTs from the new inverted-$F0$ conditions did not converge with the other conditions in reverberation. Tukey pairwise comparisons confirmed that all three $F0$ contours differed from each other in anechoic conditions (normally intonated versus monotonous, $q=7.01$, $p<0.001$; normally intonated versus inverted, $q=12.16$, $p<0.001$; monotonous versus inverted, $q=5.15$, $p<0.01$), but that in reverberant conditions the inverted condition produced higher SRTs (intonated versus inverted, $q=9.56$, $p<0.001$; monotonous versus inverted, $q=8.51$, $p<0.001$), while the monotonous and normally intonated conditions were indistinguishable. It is worth noting that the difference between the normally intonated and the inverted conditions contracted only marginally from 2.99 dB in anechoic conditions to 2.35 dB in the reverberant conditions. Thus, the $F0$-contour×environment interaction was produced by a change in the SRTs for the monotonous conditions *relative* to the other two when the environment is changed from anechoic to reverberant.

D. Discussion

Surprisingly, the inverted-$F0$ speech was *less* intelligible than the monotonized speech, despite the fact that it sounded considerably more acceptable, and less artificial, than the monotonized speech. The best explanation we can offer for this outcome is that when the $F0$ contour is monotonized there is a simple loss of prosodic information. Prosodic information usually comes from three sources, the $F0$ contour, the intensity contour, and the rhythm of the utterance. A monotonous $F0$ contour provides no prosodic information; listeners either disregard it, or simply perform worse due to the loss of information. In the inverted-$F0$ condition, on the other hand, the intonation contour is plausible, and listeners clearly attempt to exploit it. Since it is not the correct contour, it does not provide the correct information. Indeed, it probably disturbs listeners' normal processing by providing information that conflicts with that from the rhythmic and intensive aspects of the speech. The listeners' comprehension of the sentences is thus actively mislead.

It has previously been demonstrated that distortions of normal prosodic information can affect speech processing. For instance, Cutler and Clifton (1984) made recordings of isolated two-syllable words using a trained speaker who deliberately misplaced the primary lexical stress. Reaction times in a semantic processing task showed that listeners could process correctly intonated words more rapidly than incorrectly intonated ones. However, Cutler and Clifton's experiment and similar experiments by other authors were confounded to some extent by phonetic differences between stressed and unstressed syllables (Cutler *et al.*, 1997). Cutler *et al.* conclude that the role of lexical stress in lexical access is probably quite limited for English, because few words are distinguished by prosody alone. Using a cross-splicing technique, Cutler and Darwin (1981) showed that preceding prosodic context had a strong influence on the speed and accuracy of processing of subsequent words. By independently modulating the amplitude, timing, and $F0$ cues, Cutler (1987) showed that each cue made its own contribution to this effect, although, when intensity and $F0$ cues were inconsistent (as in experiment 2), reaction times were particularly long. In addition to these effects, it is possible that distortions of vowel intrinsic pitch are making some contribution to the deleterious effect of inverted $F0$ contours.

Regarding the original purpose of the experiment, the large difference between the intonated and inverted-$F0$ conditions shows that listeners *were* able to exploit information conveyed by the $F0$ contour in the presence of reverberation. Since this difference in thresholds is of a similar magnitude in both anechoic and reverberant conditions, it seems likely that the inverted $F0$ contour continues to actively mislead listeners in the reverberant case. This outcome clarifies the interpretation of experiment 1.; the idea that reverberation destroys listeners' ability to make use of the prosodic information in the $F0$ contours must be abandoned. In both experiments, the differences between normally intonated and monotonized speech were abolished in reverberant conditions. Since reverberation does not affect prosodic processing, then this effect must be attributed to better perceptual separation of the monotonized speech, compared to the normally intonated speech under reverberation. The more robust perceptual segregation of monotonized speech in reverberant conditions can be seen from the fact that it has a lower SRT compared to the intonated and inverted conditions in the reverberant case than it does in the anechoic case.

Finally, overall differences in mean SRT between ex-

periments 1 and 2 can be mainly attributed to the change in the set of target sentences. The differences observed here are consistent with unpublished measurements by Zurek (1996) using the same recordings. These show that lists 1–12 from the Harvard corpus of sentences (used in experiment 1) tend to yield SRTs 2–3 dB lower than lists 40–73. Experiment 2 used lists 40–51, inclusive. More careful selection of interfering sentences in experiment 2 (so that they were always longer than the targets) may also have contributed to the higher SRTs observed in that experiment.

IV. CONCLUSIONS

The hypothesis that speaking in a monotone at reverberant cocktail parties would aid communication is not supported, because monotonous speech is intrinsically less intelligible than normally intonated speech. Nevertheless, we have shown that reverberation has a detrimental effect on listeners' ability to perceptually separate voices with normally intonated $F0$ contours. Reverberation also disrupts listeners' ability to exploit differences in the spatial location of competing voices/sounds. These two effects both degrade social communication in reverberant rooms, and should be considered when designing spaces intended for social interaction.

ACKNOWLEDGMENTS

This work was supported by Hanse-Wissenschaftskolleg (HWK) Lehmkuhlenbusch 4, 27753 Delmenhorst, Germany. Chris McGowan collected additional data for experiment 2. We are indebted to Chris Darwin, one anonymous reviewer, and Les Bernstein as editor for their thorough and thoughtful reviews. We are also indebted to Paul Boersma and David Weenink for the use of their Praat software.

[1]A 15% vocal-tract shortening is achieved by resampling to a sampling rate 15% lower, and then playing back using the original sampling rate. This operation also increases the articulation rate and the $F0$ by 15%, so the speech is first resynthesized with a time-warp so that its articulation rate is reduced by 15%. At this point, the $F0$ can also be transformed. A reduction by 15% will compensate for the resampling. However, in the present application the $F0$ was increased overall by a factor of 1.8 in order to bring it into the female range. Thus, the $F0$ transformation applied was 1.8/1.15.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

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

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

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

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am. **25**, 957–959.

Culling, J. F. (**1996**). "Signal processing software for teaching and research in psychoacoustics under UNIX and X-windows," Behav. Res. Methods Instrum. Comput. **28**, 376–382.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

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

Culling, J. F., and Darwin, C. J. (**1993b**). "The role of timbre in the segregation of simultaneous voices with intersecting $F0$ contours," Percept. Psychophys. **54**, 303–309.

Culling, J. F., Summerfield, 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**, 71–96.

Cutler, A. (**1987**). "Components of prosodic effects in speech recognition," in *Proceedings of the Eleventh International Congress of Phonetic Sciences*, Tallinn, Estonia, Vol. 1, pp. 84–87.

Cutler, A., and Clifton, C. (**1984**). "The use of prosodic information in word recognition," in *Attention and Performance X: Control of Language Processes*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ).

Cutler, A., and Darwin, C. J. (**1981**). "Phoneme-monitoring reaction time and preceding prosody: effects of stop closure duration and of fundamental frequency," Percept. Psychophys. **29**, 217–224.

Cutler, A., Dahan, D., and van Donselaar, W. (**1997**). "Prosody in the comprehension of spoken language: A literature review," Lang. Speech **40**, 141–201.

Darwin, C. J., and Culling, J. F. (**1990**). "Speech perception seen through the ear," Speech Commun. **9**, 469–476.

Darwin, C. J., and Hukin, R. W. (**2000**). "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," J. Acoust. Soc. Am. **108**, 335–342.

de Cheveigné, A. (**1997**). "Concurrent vowel identification. III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.

de Cheveigné, A., McAdams, S., and Marin, C. (**1997**). "Concurrent vowel identification. II. Effects of phase, harmonicity, and task," J. Acoust. Soc. Am. **101**, 2848–2856.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Lea, A. (**1992**). "Auditory models of vowel perception," Ph.D. thesis, Nottingham (unpublished).

O'Shaunessy, D., and Allen, J. (**1983**). "Linguistic modality effects on fundamental frequency in speech," J. Acoust. Soc. Am. **74**, 1155–1171.

Parsons, T. W. (**1976**). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. **60**, 911–918.

Peterson, P. M. (**1986**). "Simulating the response of multiple microphones to single acoustic 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 azimuth of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

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

Summerfield, Q., and Culling, J. F. (**1992**). "Periodicity of maskers not targets determines ease of perceptual segregation using differences in fundamental frequency," J. Acoust. Soc. Am. **92**, 2317 (A).

Zurek, P. (**1996**). Personal communication.