The role of fundamental frequency contours in the perception of speech against interfering speech

Christine Binns[a] and John F. Culling
Department of Psychology, University of Cardiff, Tower Building, Park Place, Cardiff, CF10 3AT, Wales

(Received 9 February 2006; revised 8 March 2007; accepted 1 June 2007)

Four experiments investigated the effect of the fundamental frequency (F0) contour on speech intelligibility against interfering sounds. Speech reception thresholds (SRTs) were measured for sentences with different manipulations of their F0 contours. These manipulations involved either reductions in F0 variation, or complete inversion of the F0 contour. Against speech-shaped noise, a flattened F0 contour had no significant impact on SRTs compared to a normal F0 contour; the mean SRT for the flattened contour was only 0.4 dB higher. The mean SRT for the inverted contour, however, was 1.3 dB higher than for the normal F0 contour. When the sentences were played against a single-talker interferer, the overall effect was greater, with a 2.0 dB difference between normal and flattened conditions, and 3.8 dB between normal and inverted. There was no effect of altering the F0 contour of the interferer, indicating that any abnormality of the F0 contour serves to reduce intelligibility of the target speech, but does not alter the masking produced by interfering speech. Low-pass filtering the F0 contour increased SRTs; elimination of frequencies between 2 and 4 Hz had the greatest effect. Filtering sentences with inverted contours did not have a significant effect on SRTs. [DOI: 10.1121/1.2751394]

PACS number(s): 43.71.Es, 43.71.Gv [MSS] Pages: 1765–1776

I. INTRODUCTION

Intonation is known to have many functions in language, including helping with syntactic disambiguation (Cutler *et al.*, 1997), distinguishing between new and given information (Cutler *et al.*, 1997), syllable stress (Lehiste, 1970), segmentation (Liss *et al.*, 1998) as well as voicing cues (Haggard *et al.*, 1981) and cues to vowel identity (Traunmüller, 1981). Less is known about the overall impact of intonation on speech intelligibility.

The following set of experiments aims to investigate the impact of intonation on speech intelligibility under different listening conditions—speech-shaped noise and a single-talker interferer. The effect will be measured using speech reception thresholds while manipulating the F0 contour, either by reducing the amount of F0 variation incrementally or by F0 inversion.

Previous experiments have explored the role of intonation in speech recognition by flattening the F0 contour and comparing the intelligibility of these utterances to normally intonated ones in a quiet listening environment, with a background of white noise or against a competing talker (Wingfield *et al.*, 1984; Assmann, 1999; Laures and Weismer, 1999). These studies showed a detrimental effect of flattening the F0 contour on sentence intelligibility.

In the study by Laures and Weismer (1999), two male speakers read nine low-predictability sentences selected from the Speech Perception in Noise test with a pronounced prosody. Each of these sentences was resynthesized using the LPC resynthesis technique. The F0 contour of the voiced segments was flattened by adjusting their F0 value to the mean F0 of the sentence. Ten listeners were presented with both normally intonated and monotonous sentences in a background of white noise and asked to transcribe them. On a second presentation of the stimuli, listeners were asked to rate the sentences for intelligibility on a scale of 1–7, with 1 being unintelligible and 7 being highly intelligible. A significant difference between the normally intonated and monotonous sentences was found for both the number of words transcribed correctly and rated intelligibility. Among the explanations proposed for these findings was that the rise and fall of the F0 contour has been found to direct the listener's attention to the content words of the utterance (Cutler and Foss, 1974); therefore without these cues, the intelligibility of the utterance is lowered.

Another contributing factor may be the presence of F0 movement. Culling and Summerfield (1995) investigated the effect of frequency modulation on the identification of concurrent vowel sounds presented concurrently with interfering vowels. Listeners were asked to report both the presentation intervals and the identity of a target vowel in conditions where the target vowel was modulated halfway through the stimulus and either none, one, or two of the interfering vowels were modulated throughout. The modulation of the target vowel was deferred since a pilot study showed that this enhances the effect of frequency modulation on vowel prominence. Results showed that F0 modulation increases the perceptual salience and identification of vowels against the background of other unmodulated vowels. These results suggest that frequency modulation can enhance the target by making it more salient than the competing sound.

Our studies aim to further investigate these effects by decreasing the amount of F0 variation gradually with the inclusion of conditions where the contour retains only half or a quarter of the total variation present in a standard contour,

[a]Electronic mail: binnsc@cf.ac.uk

along with an inverted F0 condition. An inverted F0 contour retains the variation present in a standard F0 contour, but these variations are reversed so that where there would be a rise in the normally intonated contour, there is now a fall, and vice versa. Inverted F0 contours have been found to detrimentally affect the intelligibility of an utterance by Culling *et al.* (2003) in experiments investigating the effect of reverberation on listeners' abilities to separate two competing voices. This result indicates that it is not simply the movement *per se* of the F0 contour that aids the listener in his/her comprehension of the utterance, since this is still available in the inverted F0 contour, but perhaps more that the new contour is linguistically misleading. In a normally intonated sentence, content words are highlighted by a number of cues, including their F0. Since this F0 cue is inverted in an inverted contour, it is possible that the listener's expectations of which words are important will be confused and hence sentence intelligibility will be compromised. Hillenbrand (2003) has also experimented with monotonized and inverted F0 contours. He measured the intelligibility of synthetic speech presented in quiet. When the F0 contour was flattened there was a reduction in intelligibility, but when it was inverted there was no further effect. However, when the sentences were low-pass filtered at 2 kHz to reduce their overall intelligibility before manipulating the F0 contours, he found that the effect of manipulating the contours was larger and a difference between the monotonous and inverted conditions began to emerge. These results imply that the F0 contour's contribution to speech intelligibility increases in adverse listening conditions. In our experiments, adverse listening conditions will be created through the presence of background noise in an effort to further investigate these effects.

A number of stimulus features are known to produce different effects on speech intelligibility. The choice of stimuli, both targets and interferers, for these experiments will be motivated through the discussion of these features. In terms of interferers, there are three factors that we feel need to be analyzed: the role of F0 differences, dip-listening effects, and informational masking.

First, the identification of target speech is better when it differs in F0 from the interfering speech. Previous experiments have shown that listeners are better able to segregate two speech sources if they differ in mean F0 (Brokx and Nooteboom, 1982; Bird and Darwin, 1998; Assmann, 1999; Drullman and Bronkhorst, 2004). Brokx and Nooteboom (1982), using monotonous target and interfering speech, found that word recognition rates increased from approximately 40% correct at 0 semitones difference to roughly 60% correct at 3 semitones difference. A further measurement taken at a 12 semitone difference showed a decline in the F0 segregation advantage, compared to the 3 semitone difference, with approximately 50% correct word recognition. Also using monotonized speech, Bird and Darwin (1998) have shown that the segregation advantage increases beyond the 3 semitone difference found by Brokx and Nooteboom up to 8 semitones. Assmann (1999) has replicated these findings for both intoned and monotonous speech, showing a steady increase in speech intelligibility from 0 semitone difference up to 8 semitone difference between the target and interferer. Drullman and Bronkhorst (2004) found that listeners' performance gradually improves with increasing F0 difference up to 12 semitones when the speech retains naturally modulated F0s, as opposed to the speech with fixed F0s used by Brokx and Nooteboom. In naturally modulated speech, the F0 fluctuations within the speech would mean that, although the mean F0 difference between the target and interferer is fixed, the difference at any point in time would vary around this mean. Therefore, it seems that in the case of naturally modulated speech, a greater difference in mean F0 between the target and interferer produces a larger segregation advantage for the listener due to the increased chance of the target and interferer F0 values not overlapping.

In terms of F0 segregation, de Cheveigné (1993) proposed a neural-cancellation model whereby harmonic sounds could be segregated by isolating the F0 of the interfering sounds and canceling them. Hence, according to the model, a listener's ability to identify target speech more readily when it differs in F0 from the interfering speech is due to a cancellation process that perceptually removes the interfering speech on a particular F0. In this model, there is no limit, in principle, to the number of interfering harmonic sounds at different F0's that can be perceptually canceled. However, Culling *et al.* (2005), found that when target speech was presented against two interfering speech sources, the listener was best able to identify the target speech when the interfering voices were monotonized at a different F0 to the target speech, but at the same F0 as each other. When the interfering voices were on different F0's to each other as well as to the target speech, only a minimal advantage was seen compared to when all voices were monotonized on the same F0. This implies that a cancellation mechanism exists but that it can remove sounds at only one F0, as opposed to multiple sounds at different F0's. Therefore, a single-talker interferer will produce less masking than multiple speech interferers on different F0's.

Second, continuous white noise is a more effective masker than modulated noise, which, in turn, tends to produce more masking than a single-talker interferer (Carhart *et al.*, 1969; Festen and Plomp, 1990). The difference between white noise and modulated noise is that modulated noise permits dip listening whereas white noise does not. Dip listening is where the listener is given glimpses of the target speech through the interfering noise because of the natural dips and pauses present in the speech wave form. Noise that is modulated like speech therefore inherits these fluctuations, enabling the listener to dip listen. Modulated noise still produces more masking than a single-talker interferer because the speech interferer contains F0 information.

Third, informational masking is said to occur when the effect of masking cannot be attributed to energetic masking. Decreasing target-masker similarity tends to reduce the informational masking effects of the stimulus (Festen and Plomp, 1990; Durlach *et al.*, 2003; Brungart *et al.*, 2001). Brungart *et al.* (2001) showed that using the same talker for the target as for the interferer generated more masking than using a different talker of the same sex, which, in turn, created more of a masking effect than using a talker of the

opposite sex. Informational masking tends to occur in impoverished listening conditions, with few grouping cues, such as when both voices come from the same location or are within the same F0 range. The fact that more informational masking occurs when the voices are perceived to come from the same location has been found to decrease as the number of talkers increases from 3 to 4 to 6 to 10, although the spatial difference advantage does not completely disappear (Freyman *et al.*, 2004). Hence, with one or two interfering voices, recognition performance improves when the target and interferer appear to be from different locations, but as the number of interfering talkers from separate directions is increased, this cue no longer aids recognition performance. This effect is proposed to be due to the rise of informational masking as the number of voices is initially increased to two voices, and then a consequent fall as the number of voices increases to three. As the number of talkers in the interferer increases, it is possible that they begin to mask each other and hence enable the listener to pay more attention to the target voice.

A combination of all three of these factors can be seen in experiments using multiple-talker interferers. As the number of interfering talkers is increased, the amount of masking also increases (Peissig and Kollmeier, 1996; Drullman and Bronkhorst, 2004). The more background talkers present, the greater the chance of the dips in one talker's speech being filled by another talker and the less effect of F0 difference there is. As noted earlier, informational masking is also more likely to occur in situations using two or three voice interferers (Carhart *et al.*, 1975; Freyman *et al.*, 2004), and becomes less prominent as the number of voices increases above this amount.

Given the above-mentioned factors concerning the types of interferer used, speech-shaped noise and a single talker interferer will be used to simulate the effect of listening to speech in a speech background in the following experiments. The single-talker interferer will be a different talker to the target speech, and placed outside the target's F0 range in order to minimize any effects of informational masking. Speech-shaped noise acts as a speech-type background without producing the informational masking effects that would potentially be present if multiple talkers with similar characteristics were used.

The nature of the target speech can also affect intelligibility. Notably the predictability of the speech materials is an important factor affecting the intelligibility of the utterances. Boothroyd and Nittrouer (1988) found that phonemes placed in nonsense words were harder to recognize than phonemes placed in real words. Random sequences of words (e.g., "girls white car blink") were also less easily recognized than words in grammatically correct sentences with few contextual cues (e.g., "ducks eat old tape"). These were, in turn, harder to correctly identify than words placed in grammatically correct sentences with contextual cues (e.g., "most birds can fly"). Low-predictability sentences will be used in this experiment to ensure than the listeners can make relatively little use of contextual cues to predict the words in the sentences. Due to the need for continuously voiced sentences in Experiment 4, not all sentences are from the same corpus, but they do maintain a low-level of predictability.

The following experiments measure the impact of F0 cues in the intelligibility of speech against noise. Reducing the F0 incrementally from the variation present in a normally intonated contour to a monotonous contour, with no variation, will enable us to see whether the more variation in the contour, the more intelligible the speech or whether it is the case that just a small amount of F0 variation can give the listener enough cues to correctly transcribe the target speech. If it is the case that the listener extracts information from the F0 contour to enhance speech intelligibility, placing inverted F0 contours on the speech should mislead the listener because the F0 cues will now be incorrect.

II. EXPERIMENT 1

A. Method

1. Listeners

Twenty paid participants were recruited from the Cardiff University Participation Panel. All were native speakers of English and normal hearing (self-report). None of the listeners were familiar with the sentences used in the study.

2. Stimuli

The set of sentences used in this experiment was from the Harvard IEEE corpus (Rothauser *et al.*, 1969). The recordings, made at M.I.T. of male voice DA, were digitized at 20 kHz sampling rate with 16 bit quantization. All the sentences were low predictability and included five nominated key words. For example, "a WISP of CLOUD HUNG in the BLUE AIR."

Each sentence was manipulated using the Praat PSOLA speech analysis and resynthesis package. The F0 contour of each sentence was manipulated using Eq. (1); m is the coefficient for the particular manipulation; $\overline{F0}$ is the mean fundamental frequency. The mean F0 of the target speaker was 107 Hz, with an average pitch range of 120 Hz. By inverting the F0 contour, there were times when the F0 reached below the acceptable range allowed by Praat; Praat cannot synthesize F0 values lower than about 70 Hz, hence when inversion produced values below this limit, Praat incorrectly synthesized the F0 at a higher value. For this reason, the F0 contour of each sentence was multiplied by 1.5 in order to raise the mean F0 so that the F0 contour could be inverted without falling below the permitted F0 range for the Praat software package. The F0 range of different voices is approximately equivalent on a logarithmic scale (Graddol, 1986; Traunmuller and Branderud, 1989; Nolan, 2003), hence this was considered preferential for manipulating the F0 contour across sentences to a linear scale. After this, each of the sentences was resynthesized,

$$F0' = [1.5\overline{F0}\exp(m\ln(F0/\overline{F0}))]. \quad (1)$$

Five different manipulation coefficients (m) of the F0 contour were applied (1, 0.5, 0.25, 0, and −1), corresponding to the five conditions (standard, half, quarter, monotone, and inverse), see Fig. 1. The standard contour refers to the nor-

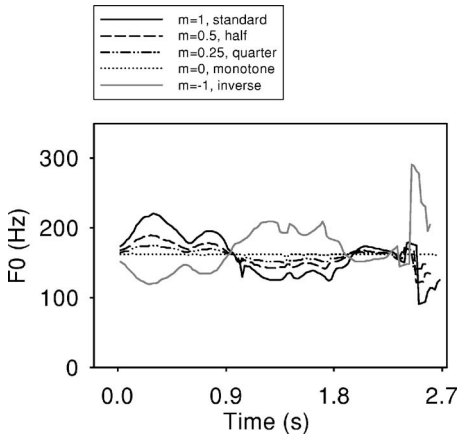

FIG. 1. Manipulations (m) of the F0 contour for an example sentence ("Greet the new guests and leave quickly") from Experiment 1. Manipulations are $m=1$, 0.5, 0.25, 0, and −1, corresponding to the five conditions of the experiment: standard, half, quarter, monotone, and inverse, respectively.

mal intonation placed upon that sentence by the speaker. The half and quarter conditions follow the same general shape of the standard contour, but the amount of variation in both is reduced; in the half condition this variation is half that of the standard contour and in the quarter, it is quarter. The monotone condition refers to a monotonized F0 contour, and the inverse to an inverted contour.

Speech-shaped noise was created by processing white noise with a digital filter whose magnitude response was equal to the long-term average spectrum of the entire set of unmanipulated target sentences. This noise was used as the interferer and edited to be, on average, 0.1 s longer than the speech targets so that no part of the speech target was unmasked at any stage.

3. Procedure

Participants were seated in an IAC single-walled sound-attenuated booth in front of a computer screen visible through the booth window. They listened to stimuli over Sennheiser HD-590 headphones and responded to them using a keyboard placed inside the booth.

Speech reception thresholds (SRTs) were measured for each condition in the experiments (Plomp and Mimpen, 1979; Culling and Colburn, 2000). In the first phase of a SRT measurement, the sound level of the target started low at –28 SNR. The listener's instructions were to attend to the voice and ignore the interfering noise. The listener was allowed to hear the same target sentence and interferer repeatedly at the start of each run by pressing the return key on the keyboard. Each time this key was pressed, the level of the target sentence was increased by 4 dB. Once the listener believed that they could hear more than half the sentence correctly, they typed out their transcript. The correct sentence was then displayed underneath the listener's transcript, containing five capitalized keywords. The listener compared these words to those in their own sentence and entered the number of key words that they had heard correctly (0–5). The second phase of the measurement then began. The sound level of the target was set according to their performance on the first sentence. Thus if the listener heard three or more words correctly, the level was 2 dB below that of the initial level set by the listener in the first phase by pressing the return key. If the listener achieved less than three words correct, the level was raised by 2 dB. The listener could now only hear the sentence once and an SRT was measured using a one-up/one-down adaptive SRT technique whereby if the listener heard three or more words correctly, the level of the target speech was decreased by 2 dB; otherwise it was increased by 2 dB. Listeners were presented with a new sentence on each trial, even if they heard none of the key words correctly. The SRT measurement was complete once the listener had heard and written out their transcript for all ten target sentences. The measured SRT for the trial was taken as the average signal-to-noise ratio for sentences three to ten.

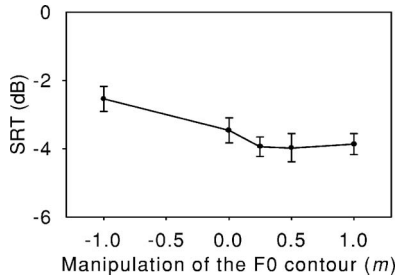

FIG. 2. Mean SRT measurements (dB) across all listeners for the different manipulations of the F0 contour (m) in Experiment 1. Manipulations $m=1$, 0.5, 0.25, 0, and −1 correspond to the following conditions: Standard, half, quarter, monotone, and inverse. Error bars represent ±1 s.d.

The session began with two practice SRT measurements with normally intonated sentences played against speech-shaped noise interferers. This practice familiarized the listener with the task. Ten further SRT measurements were made. Two SRT measurements were made for each condition (standard, half, quarter, monotone, and inverse). The order of the presentation of the conditions was rotated for each listener, while the sentence material stayed in the same order. Thus, the sentence order remained independent of the condition, with participants starting with the same sentence list, but different conditions. This ensured that each of the 100 target sentences was presented to every listener in the same order and contributed equally to each condition as well as making sure that each condition was presented in each position within the session, counterbalancing order effects.

B. Results and discussion

Figure 2 shows that there was a slight increase in SRTs as m was reduced, from the standard targets ($m=1$) to the monotonous ($m=0$) targets, with a greater increase for targets with an inverted F0 contour ($m=-1$). The difference between speech with a standard contour ($m=1$) and monotonous speech ($m=0$) was 0.4 dB and was 1.3 dB between standard speech ($m=1$) and speech with an inverted F0 contour ($m=-1$). However, the difference was smaller than the 4 dB difference between normal and inverted F0 contours found by Culling *et al.* (2003), who used speech as an interferer. A repeated measures ANOVA found a significant main effect of the F0 contour ($F(4,76)=3.525; p<0.02$). Post-hoc Tukey HSD tests, which include corrections for multiple comparisons, showed there to be a significant difference be-

tween the standard, half, and quarter conditions ($m=1,0.5,0.25$) and the inverted condition ($m=-1$) at the 0.05 significance level. No other comparisons were significant.

These results show that flattening the F0 contour causes a reduction in SRT; however, since this effect was nonsignificant, the results do not reflect those of earlier studies where flattening the F0 contour has proved significantly detrimental to speech intelligibility (e.g., Assmann, 1999; Laures and Weismer, 1999; Laures and Bunton, 2003). Inverting the contour, on the other hand, significantly reduced the intelligibility of the target speech. The fact that the quarter and half conditions did not substantially reduce the sentence intelligibility, whereas inverting the contour did, implies that as long as there is only a small amount of F0 modulation in the "correct" direction, the intelligibility of the utterance is almost unaffected.

The inverted condition produced a statistically significant decrement but still one that was smaller than observed by Culling *et al.* (2003). One difference between the current experiment and that of Culling *et al.* (2003) was that a single-talker interferer was used in their experiment instead of speech-shaped noise. Single-talker maskers have been shown to produce less masking than speech-shaped noise, which is thought to be due in part to the listener's ability to exploit the dips in the temporal envelope of the speech, which are not present in the noise (Festen and Plomp, 1990; Peissig and Kollmeier, 1997). However, the differences in fundamental frequency and intonation contours between the target and masker can be used to perceptually segregate the two talkers, hence it could be that the fundamental frequency contours are more important to sentence intelligibility against a single-talker interferer than a noise masker. For this reason, a single-talker interferer was used in the second experiment in order to see whether this larger difference could be replicated by changing the type of interferer. As seen earlier, inverting the F0 contour of the target speech detrimentally affected the intelligibility of the utterance. Culling *et al.* (2003) also manipulated the F0 contour of the speech interferer in their experiment by inverting it. Manipulating the F0 contour of the interfering speech as well as that of the target speech prevents an effect on the mean F0 difference. It is unclear whether the greater effect seen in the Culling *et al.* (2003) paper was due to the inversion of the target or the interferer. Therefore, the F0 contour of the target and the interferer were manipulated parametrically in this experiment in order to investigate whether the large effect observed by Culling *et al.* (2003) was an effect of adjusting the target, the interferer, or both.

III. EXPERIMENT 2

A. Method

1. Listeners

Eighteen undergraduate Psychology students were used as participants in this experiment. All were normally hearing native speakers of English. None of the participants were familiar with the sentences used, or had they taken part in the previous experiment.

2. Stimuli

The set of sentences used here was selected from the same corpus as in the previous experiment. However, the recordings of another male voice, CW, were used instead of DA.[1] This voice was also digitized at 20 kHz with 16 bit quantization. The sentences were once again manipulated using Praat PSOLA. The following formula was applied to the F0 contour of each sentence before resynthesis, adjusting the average F0 of all the target sentences to 125 Hz to enable a consistent 9 semitone difference to be set between the target and interfering sentences,

$$F0' = [125 \exp(m \ln(F0/\overline{F0}))]. \quad (2)$$

Only three values of m were used this time ($m=1,0,-1$) corresponding to the three conditions (standard, monotone, and inverse).

Speech interferers were used instead of noise and were created using ten sentences from the Harvard IEEE corpus. The recorded voice used was DA. The F0 contour of the interferers was manipulated using Praat PSOLA as for the target sentences. Equation (3), which results in a fixed mean F0 of 210.25 Hz, a 9 semitone difference between the target and interferers, was applied to the sentences before resynthesis,

$$F0' = [210.25 \exp(m \ln(F0/\overline{F0}))]. \quad (3)$$

Thus, the mean F0 of the interferers was set at an average of 9 semitones higher than that of the target sentences. This was done to avoid potentially confounding effects of the F0 difference (Assmann, 1999; Brokx and Nooteboom, 1982; Culling and Darwin, 1994; Bird and Darwin, 1998). For instance, if both the target and interferer were monotonized with the same F0, there would be no F0 differences to exploit, but if the target had a normal F0 contour and the interferer was monotonous, the differences in instantaneous F0 could be used to differentiate between the two at the points in time where the normal F0 contour deviates from the mean.

By introducing a constant difference in mean F0, we sought to minimize differences in mean F0 difference between the various conditions of contour manipulation. Bird and Darwin (1998) showed there to be a steady increase in sentence identification from 0 to 8 semitone F0 difference between the target and masker. In the experiment by Brokx and Nooteboom (1982), the advantage of the F0 difference was found to decrease at one octave (12 semitones). However, no research has indicated what happens between 8 and 12 semitones using monotonized F0s. The 9 semitone difference used here was felt to be as large a difference as could be employed without encountering the decline in the F0 difference advantage found by Brokx and Nooteboom (1982), while at the same time being large enough to reduce the chance of overlap of the F0 contours. This ensured that any effect found from manipulating the contour was due to the contour change itself and not the difference in F0 between the target and interferer. Figure 3 shows the different manipulations of both the target and interferer contours, around their F0 means.

Nine conditions were set up, using each possible combination ($m=1,0,-1$) of both the target and interfering F0

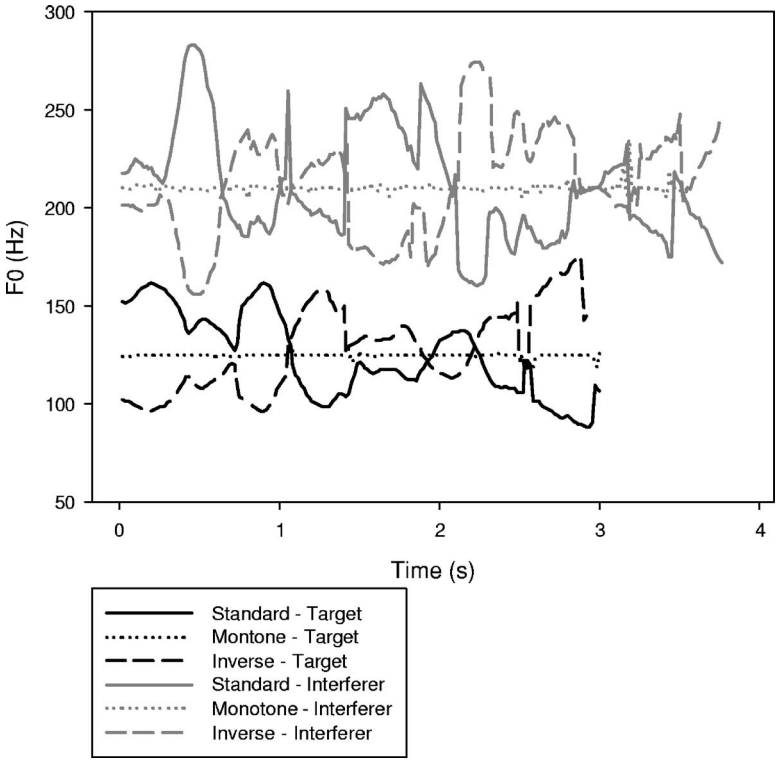

FIG. 3. Manipulations (m) of the F0 contour for both the target and interferer sentences in Experiment 2. The target speech is "Where were they when the noise started?" and the interfering sentence is "A ridge on a smooth surface is a bump or flaw." The interfering sentence in this case is 0.6 s longer than the target sentence. The mean F0 of the target is set to 125 dB, with the mean of the interferer being 9 semitones higher, at 210.25 dB. Manipulations are $m=1$, 0, and −1 for both the target and interfering contours, corresponding to the standard, monotone, and inverse conditions, respectively.

contour. Hence the target F0 contour was manipulated in three ways to create a standard, monotonous, and inverse condition for the target. The interferer F0 contour was also manipulated to be standard, monotonous, or inverse.

3. Procedure

The procedure was similar to that in Experiment 1, measuring SRTs for each of the conditions within the experiment. Each experiment, therefore, consisted of nine runs, each with ten sentences, along with a practice run at the start of the experiment to familiarize the listener with the task.

As in the previous experiment, at the start of each run, the target level would be set low at –28 SNR, whereas the interferer level was set high. The interfering sentence remained the same in both content and manipulation throughout each run. However, in order to differentiate the target from the interferer, the listener was instructed to pay attention to the quieter sentence (target) at the start of the run and ignore the louder sentence (interferer). The listener typed out their transcript, as in Experiment 1, as soon as they judged they could hear more than half of the target sentence and scored the number of words they heard correctly. SRTs were recorded for each condition. Once again, the order of the conditions was rotated for each listener, while the sentence material stayed in the same order.

B. Results and discussion

Figure 4 shows that there was a larger difference in SRT between the normally intonated, monotonous, and inverted target speech than in Experiment 1, where the interferer was speech-shaped noise. The difference between normally intonated speech and monotonous speech was 2.0 dB, and was 3.8 dB between normally intonated speech and speech with an inverted F0 contour.

A repeated measures ANOVA found a significant main effect of the target F0 contour ($F(2,24)=18.288$; $p<0.001$). Post-hoc Tukey HSD tests showed there to be a significant difference between the standard ($m=1$) target F0 condition and inverse ($m=-1$) condition ($p<0.01$). The difference between the monotone ($m=0$) and inverse ($m=-1$) conditions, and the monotone ($m=0$) and standard ($m=1$) conditions was also significant ($p<0.05$). There was no effect of varying the F0 contour of the interferer, indicating that the inversion effect is not mediated by the interferer.

Analyses of the listener's transcripts showed no errors from intrusions from the interfering speech. Instead, errors were unrelated errors in the target sentences, such as mis-

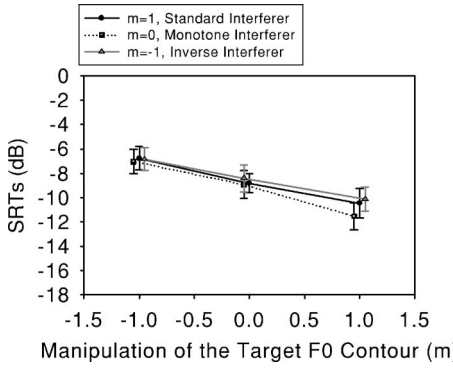

FIG. 4. Mean SRT measurements across all listeners for each of the nine conditions in Experiment 2. The three lines represent the manipulations ($m=1$, 0, −1; standard, monotone, and inverse) of the F0 contour for the interferer, with the x axis representing manipulations ($m=1$, 0, −1; standard, monotone, and inverse) of the target F0 contour. Error bars represent ±1 s.d.

heard and missing words. The implication of this is that the listeners were better able to follow and interpret the target speech where they had an increased amount of correct F0 information, rather than becoming confused by the content of the interfering speech.

As mentioned earlier, the fact that there was a larger difference between the conditions in this experiment than in Experiment 1 implies that there is something about a speech background that requires the listener to rely more heavily on the F0 contour of the target sentence in order to understand it. Mattys (2004) has also found that different cues are used in different circumstances. He found that coarticulation cues were more prominent than stress cues in quiet, whereas stress cues became more important than coarticulation cues when the speech was presented in a white noise background. Hence, it could be the case that cues that do not normally influence speech intelligibility greatly in certain conditions become more important as the conditions change. In the same way as the stress cues becoming more salient in white noise compared to quiet, cues from the F0 contour may have a greater impact on speech intelligibility in a background of speech, where the listener is required to select and follow one of two voices, rather than in a speech-shaped noise background.

Having established that a robust effect of the F0 contour is observed when the target is manipulated and presented against a speech interferer, the next experiment returned to the manipulations performed in the first experiment, where the F0 variation was reduced systematically, hence it included the half and quarter conditions used previously. These conditions were repeated, but this time against interfering speech.

IV. EXPERIMENT 3

A. Method

1. Listeners

Ten paid participants were recruited from the Cardiff University Participation Panel. All were normally hearing native speakers of English. None of the listeners were familiar with the sentences used in the study or had they taken part in either of the previous experiments.

2. Stimuli

The target sentences used here were from the same corpus as in Experiment 2, using the same speaker, CW. The interferer sentences were spoken by DA, as in Experiment 2. The same equations as used in Experiment 2 were applied to the target and interfering speech in order to achieve the different F0 means, 125 and 210.25 Hz, respectively. The interferer remained normally intonated ($m=1$), since altering the F0 contour of the interferer did not make a difference in Experiment 2. The same F0 modulations as used in Experiment 1 were employed for the target speech ($m=1, 0.5, 0.25, 0, -1$) to represent the five different conditions (standard, half, quarter, monotone, and inverse) (see Fig. 5). The same procedure was followed as in Experiment 2.

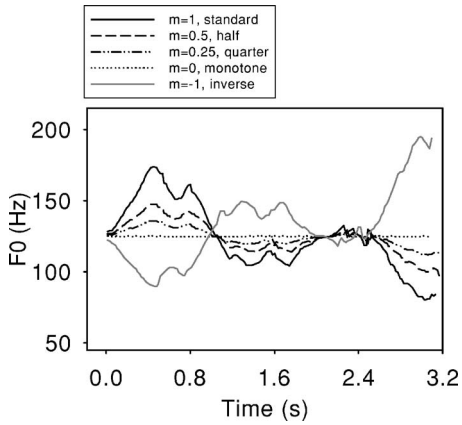

FIG. 5. Manipulations (m) of the F0 contour for an example of the target speech from Experiment 3. The sentence corresponding to the F0 contour is "All sat frozen and watched the screen." Manipulations are $m=1$, 0.5, 0.25, 0, and −1, corresponding to the five conditions of the experiment: Standard, half, quarter, monotone, and inverse, respectively.

B. Results and discussion

Figure 6 shows that there were again larger differences between each of the conditions than in Experiment 1. This confirms the results from Experiment 2 that using speech as the interferer increases the importance of the F0 contour in understanding the target speech.

No difference was found between the standard and half conditions. There was a 1.4 dB difference between the half and the quarter conditions, a further 0.2 dB difference between the quarter and the monotone conditions, and a 0.9-dB difference between the monotone and inverted conditions. A repeated-measures ANOVA showed a significant main effect of the F0 contour ($F(4,36)=3.722$; $p<0.05$). Post-hoc Tukey HSD tests showed a significant difference ($p<0.05$) between the standard ($m=1$) and inverse condition ($m=-1$), and the half ($m=0.5$) and inverse condition ($m=-1$).

The significant effect of F0 modulation implies that it is easier for the listener to separate and understand the target sentence from the single-talker interferer if the target F0 contour follows an appropriate F0 pattern. The lack of difference in decibels between the standard ($m=1$) and half ($m=0.5$) conditions implies that F0 variation can be reduced to a certain degree without hindering the intelligibility of the utter-

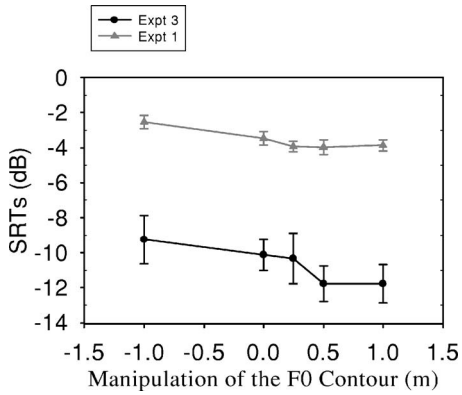

FIG. 6. Mean SRT measurements (dB) for the different manipulations ($m=1$, 0.5, 0.25, 0, −1; standard, half, quarter, monotone, and inverse) of the F0 contour in both Experiments 1 and 3. Error bars represent ±1 s.d.

ance. The difference in SRT levels between the standard ($m=1$) and quarter ($m=0.25$) conditions has become more marked than in Experiment 1, but remains insignificant. This larger difference, combined with the very slight difference between the quarter ($m=0.25$) and monotone ($m=0$) conditions, indicates that a small amount of variation in the F0 contour does not aid the listener much above the monotone. Hence it seems that, although reducing the amount of F0 variation has a progressively detrimental effect on speech intelligibility, a stronger effect is seen with F0 inversion, indicating that it is more important to have a "correct" F0 contour highlighting the important syllables in an utterance than having F0 variation.

V. EXPERIMENT 4

The previous experiments have shown that reducing variation in the F0 contour, and inverting the F0 contour cause detrimental effects to sentence intelligibility. What has not been shown by these experiments is where the important modulation frequencies for this effect lie. Low-pass filtering removes the high-frequency components of the F0 contour while retaining the low-frequency components. By progressively removing the higher frequencies, the frequencies most important for intelligibility can be revealed. Thus, rather than acting as a means for comparison with Experiments 1–3, this experiment will contribute to their results providing information on the whereabouts of frequencies involved in the F0 inversion effect.

Low-pass filtering the amplitude modulation spectrum of speech has shown that the important frequencies for the modulation of the speech signal as a whole lie between 4 and 16 Hz (Drullman *et al.*, 1994), with the most important frequency being 4 Hz, the syllable rate of speech. English is a stress-timed language (Cruttenden, 1986), which means it involves the rhythmic alternation of stressed and unstressed syllables. Therefore, the shape of the F0 contour is dependent on the accents placed on these syllables. For this reason, it would seem likely that the most important modulation frequencies for the F0 contour will lie around the syllable rate of speech, 4 Hz. The following experiment will test this prediction through low-pass filtering the F0 contour of the speech.

A. Method

1. Listeners

Twenty paid participants were recruited from the Cardiff University Participation Panel. All were normally hearing native speakers of English. None of the listeners were familiar with the sentences used in this study or had they taken part in Experiments 1–3.

2. Stimuli

In order to low-pass filter an F0 contour, it needs to be an uninterrupted wave form. For the F0 contour to be uninterrupted, the speech used needs to be continuously voiced. For this reason, 100 continuously voiced sentences were created. The key words were selected from the MRC psycholinguistic database to contain only vowels, nasals, liquids and voiced fricatives, and on the basis of their Francis and Kucera (1982) frequency being under 1000. The sentences were written in the style of the IEEE sentences in that they each contained five key words by which the listeners would rate their responses and using connecting words that were also fully voiced. Due to the limitations on the sentence construction for these stimuli, that they need to be continuously voiced, the sentences tended to be less contextually coherent than the Harvard sentences, although this was not tested. The fewer contextual cues present, the harder it is for the listener to follow the sentences (Boothroyd and Nittrouer, 1988), hence SRT values were expected to be higher for these stimuli. The list of continuously voiced sentences can be found in the Appendix.

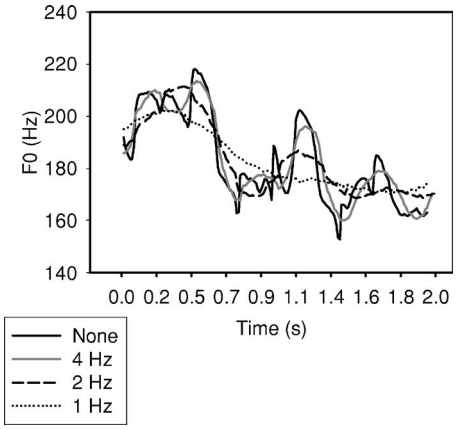

FIG. 7. Low-pass filtered F0 contours compared to an unfiltered F0 contour for the same continuously voiced sentence from Experiment 4. The sentence is "The raven rose over the rim of the ravine." The contour is filtered at 1, 2, and 4 Hz.

The sentences were then recorded using a Sennheiser K6 microphone with an ME62 omnidirectional capsule. The signals were conditioned and digitized using Tucker Davis Technologies equipment (an MA1 microphone amplifier and a DD1 digital-to-analogue converter) at 20 kHz with 16 bit quantization. The F0 contours of the sentences were extracted and low-pass filtered at cut-off frequencies of 1, 2, and 4 Hz (see Fig. 7), and, if required, inverted, before resynthesis using the Praat PSOLA speech analysis and resynthesis package as in the previous experiments.

The following formula was applied before resynthesis, setting all the sentences at an average F0 of 210 Hz,

$$F0' = [210 \exp(m \ln(F0/\overline{F0}))]. \quad (4)$$

In this experiment m was set to either 1 or −1, so that both the standard and inverse conditions could be compared when low-pass filtered.

Speech interferers were used once again and were created using ten sentences from the Harvard IEEE corpus. The recorded voice used was DA's. The following formula was applied to the interferers, setting their average F0 to 125 Hz, hence allowing a 9 semitone difference between the target and interferer contours, as in both Experiments 2 and 3, the target F0 was placed higher than the interferer F0. This was because the different target voice used in Experiment 4 had a naturally highter F0 than the voices in earlier experiments.

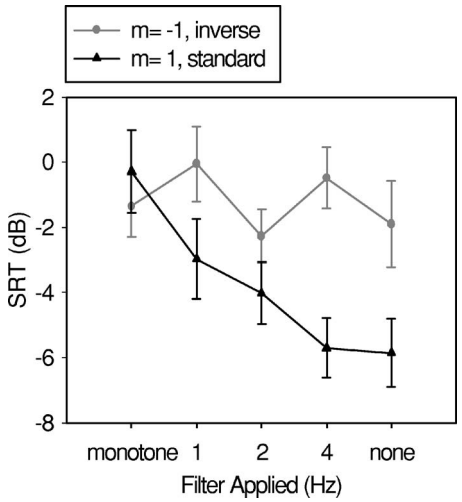

FIG. 8. Mean SRT measurements (dB) across all listeners in Experiment 4. The black line corresponds to results for the manipulated normally intonated ($m=1$) sentences, whereas the gray line represents results for the sentences with an inverted F0 contour ($m=-1$). The low-pass filtering applied to the sentences (1, 2, and 4 Hz), along with monotone ($m=0$) and standard ($m=1$) F0 manipulations, are on the x axis. Error bars represent ±1 s.d.

$$F0' = [125 \exp(m \ln(F0/\overline{F0}))]. \quad (5)$$

The variable m was set to 1 to allow for a standard contour.

Both normally intonated ($m=1$) and inverse ($m=-1$) conditions were compared for the target speech, in each of the five low-pass-filtering conditions (0, 1, 2, 4 Hz and unfiltered), giving ten conditions in total. The same procedure was followed as in Experiments 2 and 3.

B. Results and discussion

The detrimental effect of inverting the F0 contour detailed in Experiments 1–3 was replicated in this experiment ($F(1,19)=51.063$, $p<0.001$), supporting the conclusions drawn that having the correct F0 cues on the appropriate syllables aids speech intelligibility against an interfering voice.

Low-pass filtering the F0 contour was found to degrade the intelligibility of the speech ($F(4,76)=4.740$, $p<0.005$), Figure 8 shows that low-pass filtering the F0 contour produced a greater increase in SRT for the standard F0 contour than for the inverse F0 contour. This reflects the significant linear trend that was found between the F0 manipulation and low-pass filtering conditions ($F(1,19)=4.609$, $p<0.05$). That is, low-pass filtering detrimentally affected the normally intonated contour ($F(1,19)=13.243$, $p<0.001$), but did not have a significant effect on the inverse F0 contour ($F(1,19)=1.337$, $p>0.05$).

The most important frequencies for the normally intonated contour seem to lie between 2 and 4 Hz, with a 2.5 dB difference in SRT between these two conditions. However, enough information is retained in the 1 Hz filtered condition to improve the intelligibility of the utterance relative to the monotone condition, with a further 2 dB difference between these conditions. Pairwise comparisons showed that, for the standard contour, there was a significant difference ($p<0.02$) between the monotonous condition and hoth the standard condition and the standard contour filtered at 4 Hz.

The strongest effect of filtering the F0 contour was found when the frequencies between 2 and 4 Hz were removed. This implies that the most important frequencies for the F0 contour lie at or below the syllable rate of speech. Accents within the F0 contour are placed on syllables in the speech, and as shown in the above-mentioned experiments, when accents are misplaced on these syllables, as in the inverted F0 contour, speech intelligibility decreases. Indeed, deliberately mis-stressing syllables within English words has been reported to decrease response time in comparison to correctly stressed words (Cutler and Clifton, 1984). Thus, it seems that the syllable is an important unit in the F0 contour.

By reducing the amount of information in the inverse contour, it was expected that the intelligibility of the utterance would improve, since less misleading pitch information would be supplied to the listener. This prediction followed from the previous three experiments where the inverted F0 contour caused a greater detriment to speech intelligibility than a monotone F0 contour. Hence, as the inverted F0 contour is increasingly low-pass filtered, its shape tends toward a flattened or monotone F0 contour, implying that its effect on speech intelligibility would decrease. However, this did not occur.

An interesting point to note is that the difference between the monotone and standard conditions was larger with the continuously voiced sentences than in the previous experiment. On the other hand, the difference between the monotone and inverse conditions is no longer significant. There are two ways to interpret this result. The effect of the inverse F0 contour may have disappeared using continuously voiced materials. It could, be for instance, that part of the inversion effect is mediated by stop closures; the presence of onset and offset cues may serve to confuse the listener further when presented with the incorrect F0 cues in inverse contour, and, hence, once these cues are removed, the inverse contour causes no greater effect than the monotonous contour. Alternatively, it may be masked by some interaction between combined use of continuous voicing and complete monotonization of the F0 contour, which elevates thresholds in this particular condition. The latter suggestion is supported by fact that the difference between the inverse and normally intonated contours remained the same as in previous experiment (at around 4 dB). It could be that the lack of onset and offset cues from stop closures combined with the lack of variation in the F0 contour in the monotone condition is particularly detrimental to understanding and yielding SRTs that are similar to those produced by inversion.

Thus, it is unclear as to why the difference between the monotone and inverse conditions has decreased by using these continuously voiced materials. It is interesting to note that in the Hillenbrand (2003) experiment, where synthesized speech was used, a small difference was found between the monotone and inverse conditions only when the stimuli had been low-pass filtered to impair their intelligibility, suggesting that the choice of materials may influence observation of the effect.

VI. GENERAL DISCUSSION

A reduction of F0 modulation was found to have a detrimental effect on speech intelligibility, confirming previous findings (Laures and Weismer, 1999; Assmann, 1999; Laures and Bunton, 2003; Culling *et al.*, 2003). Even a small amount of F0 modulation ($m=0.25$) increased the intelligibility of the sentences against a single-talker interferer, indicating that it is the contour shape that is important to speech recognition. The fact that an inverted F0 contour caused a greater deficit than a monotonous contour implies that it is perhaps false information within the contour caused by its inversion that reduces its intelligibility over that of removing the variation from the contour through monotonization. There are a number of related ways to account for these effects: the incorrect cuing of content words, the disruption of lexical access, and the F0 contour shape in general.

A. Disrupted cuing of accented words

In a normally intonated sentence, important content words will be accented, tending to be above the average F0 of the sentence. This factor, along with the content words generally being louder and articulated more slowly, contribute to making these words acoustically clearer than the surrounding words (Lehiste, 1970). In a monotonous sentence, none of the words in the sentence are accented and all are at the same F0; therefore, there are fewer cues as to the whereabouts of the content words. In a sentence with an inverted F0 contour, however, the accented content words will now be accented in the opposite direction, for instance, a rise will become a fall and vice versa. Equally, words that were previously above the average F0 will now be below. Therefore, whereas in a monotonous sentence no F0 cues will highlight important words, in a sentence with an inverted contour, these F0 cues will be misleading and highlight words that are not particularly important to the meaning of the sentence. Not only are accented words acoustically clearer, but listeners have also been shown to use the F0 contour in order to anticipate upcoming accented words within a sentence (Cutler, 1976). For these reasons, a sentence with an inverted contour would be found to be less intelligible than a monotonous contour due to its misleading accents.

Supporting this argument is the fact that when speech is degraded, the F0 contours are relied on more heavily. Hillenbrand (2003) found that when the speech had been low-pass filtered at 2 kHz, an intelligibility difference was found between the monotone and inverted F0 conditions, implying that the misleading information in the inverted F0 contour had a greater effect when the words lost some clarity and, hence, that the F0 contour is relied upon more heavily as speech quality becomes corrupted.

B. Disruption of lexical access

Low-pass filtering the standard contour indicated that the most important frequencies for a normally intonated sentence lie between 2 and 4 Hz; when these frequencies were removed from the contour, the SRTs increased the most, corresponding to a decrease in sentence intelligibility. These frequencies are at or below the syllable rate of the sentence. Since English is a stress-timed language (Cruttenden, 1986), it involves the rhythmic alternation of stressed and unstressed syllables. Thus, it makes sense that the syllable rate is important to the intonation contour of the sentence, since it is the syllables that are assigned the particular accents. This particular result therefore implies that the clarity of the accents placed on the syllables within the sentence is important for speech intelligibility and reinforces the idea that listeners actively search for the accents within the sentence, and in doing so look to the syllables for these accents.

C. Importance of contour shape

Filtering the F0 contour also showed that a reduction of information in the normally intonated contour decreases its intelligibility, but even the lowest frequency components of a contour in the "right" direction (1 Hz filter) improved the intelligibility relative to the monotone condition. This supports the conclusion from the previous experiments implying that the contour shape is important to speech intelligibility. As mentioned earlier, the focus (accented word) of the sentence is searched for by the subject in order to quickly gain an understanding of the utterance. The F0 contour surrounding the focused word directs the listener's attention to that particular word (Cutler, 1976; Cutler and Fodor, 1979). Therefore, where there is an F0 contour in the right direction, the focused words may remain focused, to some degree, and the surrounding contour still guides the listener toward those words more readily than in a monotone condition where the variation has been removed. Hence, there are two cues present in a contour in the normally intonated condition, the accent and the surrounding contour, guiding the listener to the content words, which are not present in the monotone condition and are misleading in the inverted condition. This potentially explains why an F0 contour filtered at the 1 Hz level is more intelligible than a monotonous contour despite not containing much variation. Low-pass filtering the inverse F0 contour of sentences did not affect the intelligibility. It is not clear whether the lack of a difference between the monotone and inverse conditions is due to a decreased inversion effect or an increased monotone effect with the continuously voiced sentences.

Overall, it seems that the F0 contour is an important factor in speech intelligibility for degraded speech. Reducing the amount of F0 variation through monotonization decreases the intelligibility of speech. An inverted F0 contour causes a greater deficit than a monotonous contour, implying that it perhaps contains false information, which depresses its intelligibility. These results confirm findings from Cutler (1976) that there are two cues present in an F0 contour in the normally intonated condition, the accent and the surrounding contour guiding the listener to the content words, which are not present in the monotonous condition and are misleading in the inverted condition.

APPENDIX

the YELLOW LION WORE an IRON MUZZLE
EVERY MAN WON a LEMON RAZOR
the OILY RIVER RAN in the RURAL VALLEY

the WEARY WOMAN LAY on the MAIN LAWN
the MEAN ARMY WON EVERY WAR
a LOVELY WELL is NEAR the WOOL MILL
YOUR LAWYER is AWARE of the VALUE of the VAN
MOVE the LIME WIRE over the NEW RAIL
NINE MEN OWE ME MONEY
AIM the ARROW OVER the LOW WALL
ALLOW the LOYAL ANIMAL a WARM LAIR
the ENEMY is AWARE of YOUR INNER ZONE
WARM RAIN RAN over the IVORY MIRROR
MANY WOMEN LIVE a MILE AWAY
WARN OUR MEN of the EARLY ALARM
NONE of THEM RAN VIA the MARINA
EVEN the LIVELY EARL was AWAY ILL
NEARLY ALL the NAVY was WARY of the MAYOR
the EVEN WAVES will OVERWHELM the LONELY MARINER
ROYAL LOONIES were ALL OVER the MOOR
ELEVEN REMOVAL MEN were in the LOWER ROOM
the WARNING of NAVAL INVASION was a MARVELOUS RUSE
an ARRAY of RAW MELON is ALWAYS on the MENU
WEAVE your LINEN on the LARGE NARROW LOOM
the EVIL VILLAIN ALWAYS LIES and is EVASIVE
WOMEN RARELY MARRY in a YELLOW VEIL
the RAVEN ROSE OVER the RIM of the RAVINE
the RARE RHYTHM was the ENEMY of ALL REASON
LAZY LIMBS LAY OVER the WEIR
the WILLOWS WAVE OVER the LEISURELY LAKE
RAISE the ALARM WHENEVER the LEVEL RISES
a NORMAL MALE loses NINE ENAMEL MOLARS
in the ARENA, ROMANS EARN MANY ENEMIES
we MOVE AWAY the REVELERS in the MAUVE ROOM
SMALL MARINE MAMMALS were EVERYWHERE in the MUSEUM
we ALL WOVE WARM WOOLIES in the MILL
we KNEW of the ANIMAL'S LOW EYES and EARS
ALL LOVE UNRULY REVELRY in the EVENING
their ZEAL is ONLY the ILLUSION of EARLY RISING
MOTHER was NEVER in the VILLA with OTHER WOMEN
MINERAL EROSION REMAINS the MAIN WORRY
you will NEVER LEAN EASILY on a LAME LIMB
my MEMOIRS are VENOMOUS in EVERY LINE or WORD
MEASLES are MAINLY a MALAISE of the VERY YOUNG
the REASON WHY we were ALUMNAE was ALWAYS UNKNOWN
REAL MAYONNAISE is LOVELY or MELLOW on the NOSE
ANYONE WHO is EARLY will VIEW the MARE
the NEWS of the RALLY INVOLVES NEARLY EVERYONE
the NEW MOVIE REVEALS our ONLY ERROR
REMOVE the RIVAL MILLER ALIVE in the MORNING
LONELY MEN ROAM EVERY REALM
we OIL the REAR RAILWAY ALL YEAR
EVEN a MINIMAL WIN ALARMS MANY
EVERY VOLUME VARIES in a NOVEL WAY
our LONE MEMORY of ONE WARREN was VIVID
the MILLIONAIRE'S MANOR is OVER the NORMAN RIVER
ALL WHO OWN NAVY are NORMAL
we LEARN NEARLY ALL our MANNERS in YOUTH
NORMALLY the LEAN ROWERS WEAR LEATHER
RAW LIVER as a MEAL WORRIES ME
my MORAL LOVER is UNAWARE of the MINER'S NAME
the NERVE in your ARM NEVER ANNOYS YOU
RELY on the WARY EYE of the VAIN ALIEN
you VALUE YOUR REVOLVER MORE than ME
MOREOVER the IRONY of the NOVEL was VERY MINIMAL
RAISE the OLIVE LEVER NEAR YOU
the MINI WALL MURAL is OVER THERE
THESE ROSES LINE the AVENUE EVENLY
NO ONE KNOWS the NUN is REALLY a LIAR
we will LOSE YOU in the ZOO'S NARROW MAZE
the NEW ROYALS REIGN EVILY over the REALM
ZOOM in ANYWHERE on the NEW MOON NOW
MIRROR MY EVERY MOVE in the ROOM
the MEN'S MORALE was NEVER a WORRY of MINE
HAVE the ANNUAL REUNION in LOVELY VIENNA
they will MOAN if you REMOVE THEIR LEISURE HOUR
EARLIER the LEAVES were on the REALLY WORN LANE
USE the NEON NAIL in THEIR URN
they will MOURN the LOSS of ONE as MERRY as YOU
NEVER KNEEL on ANY WALL in the VENUE
WEIGH the ONIONS THERE NEAR the VEAL
the LOAN will RELIEVE ALL YOUR MONEY WORRIES
SLOWLY ROLL the WHEEL THROUGH the RYE
the WEATHER is USUALLY VERY RAINY in ROME
ILL ANIMALS WALLOW or ROLL in MUD
the WILL of MY MUM AMUSES MANY
i RUN MORE MILES on a MEAL of MUESLI
RENEW your VOWS of LOVE in EARLY MAY
the RAVENOUS LLAMA MAULS ALL in VIEW
in the VILE RAIN, our ONLY EWE was MAIMED
the MELLOW ROAR of the RIVER LULLS US
the ROW OVER the NEW MALL is WEAK
their AIM MERELY NUMBS MY KNEE
LAYERS of OIL OOZE over the NEW ROAD
the MEN YELL as the RAM RUINS the MINE

the ARRIVAL of the MAIL ALWAYS LEAVES them LOW

a RISE in EARNINGS will EASE THEIR WOE
you are NEVER ILL WHEN YOU are IMMUNE
ONE WOMAN'S ROLE was VERY ONEROUS
AIR YOUR VIEWS ORALLY in the UNION

[1]Speakers were changed from Experiment 1 to Experiment 2 due to the change in interferer type. In Experiment 2, a single-talker interferer needed to be used. The interferer needed to be placed at a substantially different F0 from the target speech. DA's voice sounded more natural than CW's when the mean F0 was adjusted to allow for this difference.

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

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, Q. Summerfield, and R. Meddis, Whurr, London, pp. 263–269.

Boothroyd, A., and Nittrouer S., (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

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

Brungart, D. S., Simpson, B. D., Ericscon, M. A., and Scott, K. R., (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers." J. Acoust. Soc. Am. **110**, 2527–2538.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Carhart, R., Tillman, T. W., and Greetis, E. S., (**1969**).

Cruttenden, A. (**1986**). *Intonation* (Cambridge University Press, New York).

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

Culling, J. F., and Summerfield, Q. S. (**1995**). "Perceptual segregation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay." J. Acoust. Soc. Am. **98**(2), 785–797.

Culling, J. F., Hodder, K. I., and Toh, C. Y. (**2003**). "Effects of reverberation on perceptual segregation of competing voices," J. Acoust. Soc. Am. **114**, 2871–2876.

Culling, J. F., Linsmith, G. M., and Caller, T. L. (**2005**). Evidence for a cancellation mechanism in perceptual segregation by differences in fundamental frequency." J. Acoust. Soc. Am. **117**(4), 2600.

Cutler, A. (**1976**). "Phoneme-monitoring reaction times as a function of preceding intonation contour," Percept. Psychophys. **20**, 55–60.

Cutler, A., and Clifton, C. E. (**1984**). "The use of prosodic information in word recognition," in *Control of Language Processes*, Attention and Performance Vol. **X**, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ) pp. 183–196.

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

Cutler, A., and Fodor, J. A. (**1979**). "Semantic focus and sentence comprehension," Cognition **7**, 49–59.

Cutler, A., and Foss, D. J. (**1977**). "On the role of sentence stress in sentence processing." Language and Speech, **20**, 1–10.

de Cheveigné, A. (**1993**). "Separation of concurrent vowel identification harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.

Drullman, R., and Bronkhorst, A. W. (**2004**). "Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers," J. Acoust. Soc. Am. **116**, 3090–3098.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Durlach N. I. Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, S., and Kidd, G. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity." J. Acoust. Soc. Am. **114**(1), 368–379.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Francis, W. N., and Kucera, H. (**1982**). *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton-Mifflin, Boston).

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Graddol, D. (**1986**). "Discourse specific pitch behaviour." in *Intonation in Discourse*, edited by C. Johns-Lewis (Croom Helm, London), pp. 221–237.

Haggard, M., Summerfield, Q., and Roberts, M. (**1981**). "Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F0 cues in the voiced-voiceless distinction," J. Phonetics **9**, 49–62.

Hillenbrand, J. M. (**2003**). "Some effects of intonation contour on sentence intelligibility," J. Acoust. Soc. Am. **114**, 2338.

Laures, J. S., and Bunton, K. (**2003**). "Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions," J. Appl. Photogr. Eng. **36**, 449–464.

Laures, J. S., and Weismer, G. (**1999**). "The effects of a flattened fundamental frequency on intelligibility at the sentence level," J. Speech Lang. Hear. Res. **42**, 1148–1156.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Liss, J., Spitzer, S., Caviness, J., Adler, C., and Edwards, B. (**1998**). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthritic speech," J. Acoust. Soc. Am. **104**, 2457–2466.

Mattys, S. L. (**2004**). "Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation," J. Exp. Psychol. Hum. Percept. Perform. **30**, 297–408.

Nolan, F. (**2003**). "Intonational equivalence: An experimental evaluation of pitch scales," Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 771–774.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Plomp, R., and Mimpen, A. M. (**1979**). "Speech-reception thresholds for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66**, 1333–1342.

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." I.E.E.E. Trans. Audio Electroacoust., **17**, 227–246.

Traunmüller, H. (**1981**). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.

Traunmüller, H., and Branderud, P. (1989). "Paralinguistic speech signal transformations," STL-QPSR, pp. 63–68.

Wingfield, A., Lombardi, L., and Sokol, S. (**1984**). "Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation," J. Speech Hear. Res. **27**, 128–134.