Low-frequency speech cues and simulated electric-acoustic hearing

Christopher A. Brown and Sid P. Bacon
Department of Speech and Hearing Science, Psychoacoustics Laboratory, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 26 December 2007; revised 8 December 2008; accepted 9 December 2008)

The addition of low-frequency acoustic information to real or simulated electric stimulation (so-called electric-acoustic stimulation or EAS) often results in large improvements in intelligibility, particularly in competing backgrounds. This may reflect the availability of fundamental frequency (F0) information in the acoustic region. The contributions of F0 and the amplitude envelope (as well as voicing) of speech to simulated EAS was examined by replacing the low-frequency speech with a tone that was modulated in frequency to track the F0 of the speech, in amplitude with the envelope of the low-frequency speech, or both. A four-channel vocoder simulated electric hearing. Significant benefit over vocoder alone was observed with the addition of a tone carrying F0 or envelope cues, and both cues combined typically provided significantly more benefit than either alone. The intelligibility improvement over vocoder was between 24 and 57 percentage points, and was unaffected by the presence of a tone carrying these cues from a background talker. These results confirm the importance of the F0 of target speech for EAS (in simulation). They indicate that significant benefit can be provided by a tone carrying F0 and amplitude envelope cues. The results support a glimpsing account of EAS and argue against segregation.
 [DOI: 10.1121/1.3068441]

PACS number(s): 43.66.Ts, 43.71.Ky [BCM] Pages: 1658–1665

I. INTRODUCTION

Recently, individuals with residual hearing restricted to the low frequencies (below about 500–750 Hz) have been implanted with a relatively short electrode array designed to preserve as much of the residual hearing as possible in the apical region (Gantz *et al.*, 2005; Gantz and Turner, 2003, 2004; Turner *et al.*, 2004; von Ilberg *et al.*, 1999). These individuals, in addition to full-insertion implant users who have some residual hearing in the nonimplanted ear, have the potential to combine the electric and acoustic sources of information. For both simulated and real implant processing, the addition of low-frequency acoustic stimulation often enhances speech understanding, particularly when listening to speech in the presence of competing speech (Dorman *et al.*, 2005; Kong *et al.*, 2005; Turner *et al.*, 2004). The benefit of this so-called electric-acoustic stimulation (EAS) occurs even when the acoustic stimulation alone provides little or no intelligibility (i.e., no words correctly identified).

Although little is known about the auditory processing or the acoustic cues underlying this effect, some (e.g., Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006) have suggested that listeners combine the relatively weak pitch information conveyed by the electric stimulation with the stronger pitch cue from the target talker's fundamental frequency (F0) or voice pitch in the low-frequency acoustic region to segregate target and background. It has been thought for some time that F0 aids in the segregation of competing talkers (e.g., Assmann, 1999; Assmann and Summerfield, 1990; Bird and Darwin, 1997; Brokx and Nooteboom, 1982; Culling and Darwin, 1993). Recent reports (Chang *et al.*, 2006; Qin and Oxenham, 2006) have shown indirectly that F0 is likely to play an important role independent of any role that the first formant may play. For example, the addition to vocoder stimulation of 300 Hz low-pass speech, which itself should not contain much if any first formant information (Hillenbrand *et al.*, 1995) or yield any intelligibility, improved speech intelligibility in a competing background (Chang *et al.*, 2006; Qin and Oxenham, 2006).

However, the question remains of what low-frequency cues are responsible for the EAS effect. Kong and Carlyon (2007) simulated EAS conditions, and found that voicing and amplitude envelope information provided benefit over vocoder alone. On the other hand, F0 cues provided no additional benefit at any SNR tested. They argued against F0 as a cue for segregation, and suggested that, in addition to the voicing cue, the amplitude envelope may help listeners by indicating when to listen or "glimpse" the target.

While several papers have suggested F0 as a cue for EAS, the supporting evidence has been relatively circumstantial. The primary goal of the present study was to evaluate directly the importance of F0 for EAS. A secondary goal was to evaluate the importance of the amplitude envelope of the acoustic stimulus in the EAS effect, as well as the importance of combining F0 and the amplitude envelope. To do this, we replaced the target speech in the low-frequency region with a tone that was modulated either in frequency to track the dynamic changes in the target talker's F0, in amplitude with the amplitude envelope of the low-pass target speech, or both in frequency and amplitude. There is evidence (Faulkner *et al.*, 1992) that this kind of processing can provide an aid to lip reading for hearing-impaired listeners, particularly those with limited frequency selectivity, though it is unclear whether it can yield an EAS benefit.

 0001-4966/2009/125(3)/1658/8/$25.00

In addition to the theoretical importance of determining the contribution to intelligibility by F0, there may be practical benefits as well: if EAS benefit can be demonstrated with a low-frequency tone carrying F0 (and/or the amplitude envelope), it is possible that hearing-impaired listeners with especially elevated low-frequency thresholds could benefit more from the tonal cue than from speech itself because the entire cue could be made audible due to the concentration of all the energy into a narrow frequency region, whereas only a portion of the broader band speech might be.

II. EXPERIMENT 1

Experiment 1 examined the contribution of the dynamic changes in F0 to the benefit in intelligibility from simulated EAS by replacing the low-pass speech with a tone that was modulated in frequency to track the changes in F0 that occur across an utterance. Because we expected the amplitude envelope of the low-pass speech to contribute to intelligibility as well, we included conditions in which a tone equal in frequency to the mean F0 of the target talker was modulated by the amplitude envelope of the low-pass target speech. An additional set of conditions combined the F0 and the envelope cues.

A. Method

1. Subjects

Data were collected from 25 (15 females, 10 males) fluent speakers of English, who ranged in age from 26 to 38 years and who were compensated either monetarily or with course credit for their time. All 25 listeners had pure-tone air-conduction thresholds ≤20 dB HL (ANSI, 1996) at octave and half-octave frequencies from 250 to 6000 Hz in the right ear, which was used exclusively.

2. Stimuli

Prior to testing, the dynamic changes in the target talker's F0 were extracted from each sentence using the YIN algorithm (de Cheveigné and Kawahara, 2002) with a 40 ms window size and 10 ms step size. In addition, the onsets and offsets of voicing in each utterance were extracted manually, with 10 ms raised-cosine ramps applied to the transitions.

Target stimuli consisted of the IEEE sentences (IEEE, 1969) produced by a female talker with a mean F0 of 184 Hz. Backgrounds were the AZBIO sentences (Spahr and Dorman, 2004) produced by a male (mean F0=92 Hz) or a female (mean F0=224 Hz) talker, four-talker babble (Auditec, 1997), or generic speech-shaped noise (low passed at 800 Hz, using a first-order Butterworth filter). The target speech began 150 ms after the onset of the background and ended 150 ms before the background offset. Two single-talker background sentences were concatenated when necessary. Prior to processing, the level of the broadband target speech was adjusted to 70 dB SPL and the rms level of the background stimuli was adjusted to achieve a +10 dB SNR, which was shown in pilot testing to produce about 30% correct in vocoder-only test conditions. This allowed sufficient room for improvement when a low-frequency cue was added.

The information conveyed by electric stimulation was simulated using a four-channel vocoder that employed sinusoidal carriers. The signal was bandpass filtered into four frequency bands. The logarithmically spaced cutoff frequencies of the contiguous vocoder bands were 750, 1234, 2031, 3342, and 5500 Hz. The envelope of each band was extracted by half-wave rectification and low-pass filtering (sixth-order Butterworth, cutoff frequency of 400 Hz or half the bandwidth, whichever was less). This envelope was used to modulate the amplitude of a tone at the arithmetic center of the band (the frequencies of the carrier tones were 992, 1633, 2687, and 4421 Hz). This thus simulates a 20 mm insertion depth, appropriate for "hybrid" EAS in which the electric and acoustic stimulation occur in the same ear.

The low-frequency region consisted of either target speech low-pass filtered at 500 Hz (tenth-order Butterworth) or a tone whose mean frequency equaled the mean F0 of the target talker for each sentence (overall mean F0=184 Hz). The tone was unmodulated (except for the modulation due to the onsets and offsets of voicing; this voicing cue was present in all of the tone conditions) or modulated either in frequency with the dynamic F0 changes in each target utterance, in amplitude with the envelope of the 500 Hz low-pass speech [obtained via half-wave rectification and low-pass filtering at 16 Hz (second-order Butterworth)] or both in frequency and amplitude. In all cases, the tone was audible only when voicing occurred, and the level of the tone was adjusted to be equal in rms to that of the 500 Hz low-pass speech. Note that, as was the case for the study by Kong and Carlyon (2007), the background was never present in the low-frequency region. This was done because it allowed a more sensitive measure of the contributions of each low-frequency cue of interest.

All processing was done digitally via software routines in MATLAB, and stimuli were presented monaurally using an Echo Gina 3G sound card (16 bit precision, 44.1 kHz sampling rate), Tucker Davis PA5 attenuators, and Sennheiser HD250 headphones.

3. Conditions

The output of the four-channel vocoder (target plus background) was either presented alone (V), combined with the 500 Hz low-pass target speech ($V/500$) or combined with a tone that was either unmodulated (except for voicing; V/T), modulated in frequency by the dynamic change in F0 (V/T_{F0}), modulated in amplitude by the envelope of the low-pass speech (V/T_{env}), or modulated in both frequency and amplitude ($V/T_{F0\text{-}env}$). In addition, the 500 Hz low-pass target speech and each of the tonal cues were presented in isolation without the vocoder stimulation.

4. Procedure

Participants were seated in a double-walled sound booth with one of two experimenters, who scored responses and controlled stimulus presentation. One experimenter was aware of the experimental details of each condition as it was presented, and one was not.[1] Responses were made verbally,

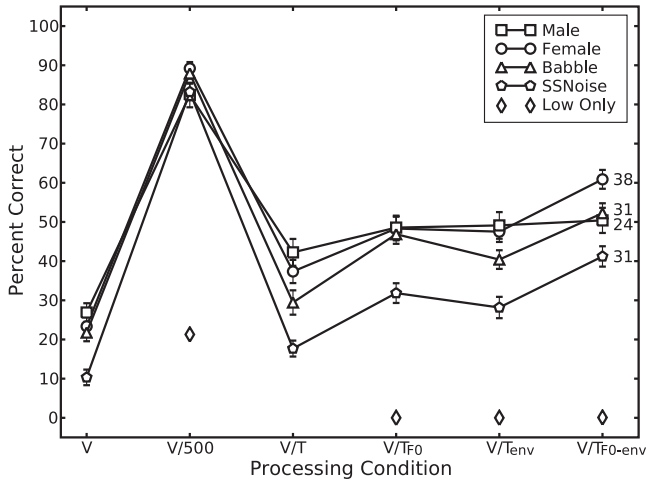

FIG. 1. Mean percent correct scores, ±1 standard error. Each curve represents a different background, which was present in the vocoder region only. Processing conditions are along the *x* axis. The output of the vocoder was presented alone (V) or with different low-frequency cues, which included 500 Hz low-pass speech ($V/500$), a tone with only voicing applied (V/T), a tone modulated in frequency by the fundamental of the target talker (V/T_{F0}), a tone modulated in amplitude by the envelope of the low-pass speech (V/T_{env}), and a tone modulated in both frequency and amplitude ($V/T_{F0\text{-}env}$). Mean intelligibility provided by the low-frequency cues themselves is depicted by diamonds. Values to the right of the $V/T_{F0\text{-}env}$ data points are percentage points of improvement over vocoder alone. Although the processing variable, which is depicted along the *x* axis, is not a continuous variable, the different levels of processing within each background are connected with lines for clarity.

and participants were instructed to repeat as much of the target sentence as they could. No feedback was provided.

Participants first heard ten unprocessed broadband target sentences presented in quiet, followed by ten sentences in quiet processed with the four-channel vocoder, to familiarize them with the target talker's voice and with the vocoder processing. Participants then heard 100 sentences of the target talker at a SNR of +10 (babble background) processed through the four-channel vocoder, combined with the low-frequency tone modulated in both frequency and amplitude ($V/T_{F0\text{-}env}$).[2]

There were 50 keywords (ten sentences) per test condition, and the presentation order of the conditions was randomized for each subject. No sentence was heard more than once.

B. Results

Figure 1 shows the mean percent correct results. Each curve represents performance in a different background, and error bars represent ±1 standard error. The different processing conditions are represented along the *x* axis. Diamonds represent performance when the respective low-frequency cue was presented alone.

A two-factor repeated-measures analysis of variance, with background and processing conditions as the main effects, revealed significant differences ($p<0.001$) within each variable. A *post hoc* Tukey analysis using a Holm–Bonferroni correction on the background variable showed all pairwise differences to be significant except male and female, and male and babble. A Tukey analysis was conducted on the different processing conditions as well; significant differences (adjusted $p<0.001$) were found between each pair of groups, except V/T_{F0} and V/T_{env}. These effects are described in more detail below.

The improvement in performance observed from the V (vocoder-only) conditions to the $V/500$ (vocoder plus 500 Hz low-pass target speech) conditions was on average, about 65 percentage points. This improvement demonstrates, in simulation, the EAS effect of combining vocoder stimulation with low-frequency speech, which itself provided only about 20% intelligibility (diamond marker at $V/500$). The improvement in performance over V in the V/T (vocoder plus tone carrying the voicing cue only) conditions averaged about 11 percentage points across backgrounds. This effect indicates that the voicing cue is informative under these conditions. The effect of the dynamic changes in $F0$ on intelligibility—above and beyond the effects of voicing—can be seen by comparing scores for V/T_{F0} with those for V/T. Across backgrounds, the improvement averaged about 13 percentage points. Similarly, the addition of a tone modulated with the envelope of the low-pass target speech to the vocoder (V/T_{env}) produced about 11 percentage points of improvement relative to V/T. Both of these differences were statistically significant ($p<0.001$). Finally, when the tone was modulated in both frequency and amplitude and combined with vocoder ($V/T_{F0\text{-}env}$), improvement over V/T averaged about 20 percentage points. Note that the tonal cues by themselves were not sufficient for any words to be reported correctly (diamond markers at V/T_{F0}, V/T_{env}, and $V/T_{F0\text{-}env}$).

For three of the four backgrounds (female talker, male talker, and speech-shaped noise), the contributions to intelligibility of $F0$ and the amplitude envelope of low-pass speech were statistically equivalent (adjusted $p>0.42$), and each cue was statistically more effective than the voicing cue alone. In addition, in these three backgrounds, the combination of F0 and amplitude envelope cues provided significant benefit over the amplitude envelope cue alone (adjusted $p<0.01$).

The amount of improvement in the $V/T_{F0\text{-}env}$ condition relative to the V condition for each background is given at the far right in Fig. 1. The largest improvement was seen in the female background (38 percentage points; circles), while the amount of improvement was 24 percentage points in the male background (squares), and about 31 percentage points in both the multitalker babble (triangles) and speech-shaped noise backgrounds (pentagons).

When the background was a male talker, nearly all of the benefit over vocoder only provided in the tone conditions was due to the voicing cue. Neither F0, nor the amplitude envelope, nor the combination of the two cues provided significantly more benefit than voicing alone (recall that the voicing cue was present in all of the tone conditions). It is unclear why this pattern of results was obtained only when the background was male.

C. Discussion

For three of the four backgrounds (female talker, babble, and speech-shaped noise), the pattern of results was similar:

F0 and the amplitude envelope of low-pass speech contributed equal, and somewhat independent and additive sources of information. In addition, each cue alone provided significant benefit over the voicing cue (recall that voicing was present in all tonal conditions), demonstrating that both F0 and the amplitude envelope are useful cues in simulated EAS conditions. These findings contrast with a recent finding (Kong and Carlyon, 2007) that showed no additional improvement in intelligibility due to the dynamic changes in F0 over the combined voicing and amplitude envelope cues.

On the other hand, performance with the male background showed a pattern of results that is more similar to that found by Kong and Carlyon (2007), in that nearly all of the improvement observed in the tone conditions could be attributed to the voicing cue, and F0 contributed no further benefit. It is unclear why this pattern of results is observed only for the male background in the current experiment and thus why the pattern of results here is generally different from that in Kong and Carlyon (2007). There are various procedural differences (e.g., sentence materials, number of vocoder channels, and carrier type in the low-frequency region) that may have contributed to the different pattern of results in the two studies. This is the focus of follow-up experiments currently under way.

1. Effects of background

Our results with speech-shaped noise are also inconsistent with those reported in the literature. Turner *et al.* (2004) showed an EAS benefit to speech intelligibility when low-frequency speech was combined with both real and simulated electric stimulations when the background was a competing talker, but not speech-shaped noise. We show a benefit under our simulated EAS conditions (compare V with $V/500$) with both types of backgrounds. This discrepancy may be due to our decision not to include the background in the low-frequency region.

Previous studies (e.g., Stickney *et al.*, 2004) have reported that a competing talker produces poorer speech intelligibility than speech-shaped noise in vocoder-alone processing. We did not obtain similar results; however, in that speech-shaped noise produced the most masking (compare different backgrounds in the V processing condition). Our results can be explained by the generic nature of our speech-shaped noise. Because it was simply low-pass filtered at 800 Hz with a first-order Butterworth filter, it had more energy in the frequency range encompassed by our vocoder (750–5500 Hz) than the speech backgrounds we used, and thus was a more effective masker.

2. Explaining the benefits of EAS

Our results show that the additional low-frequency voicing, amplitude envelope, and F0 cues can more or less independently contribute to speech intelligibility in simulated EAS. F0 has been thought for some time to aid in segregating competing talkers (e.g., Assmann, 1999; Assmann and Summerfield, 1990; Bird and Darwin, 1997; Brokx and Nooteboom, 1982; Culling and Darwin, 1993). Indeed, several recent reports (e.g., Chang *et al.*, 2006; Qin and Oxenham, 2006; Kong *et al.*, 2005) have suggested that F0 may aid in segregation of target and background in simulated EAS. However, the current results provide indirect evidence that indicate that segregation is not responsible for the benefits observed. For example, if the benefit due to F0 is explained by segregation, we might expect that the conditions which contained the greatest F0 difference between target and masker would yield the greatest benefit when target F0 information is added. However, the F0 difference between target and background was greatest when the background was male (F0 difference of 92 Hz), yet the addition of target F0 information had the least benefit with this background. Thus, the results of experiment 1 are not consistent with the segregation as an explanation for the benefits of F0 under EAS conditions.

There are other ways F0 may provide benefit as well. F0 has been shown to be important for several linguistic cues, including consonant voicing (Boothroyd *et al.*, 1988; Holt *et al.*, 2001), lexical boundaries (Spitzer *et al.*, 2007), and contextual emphasis (Fry, 1955) as well as manner (Faulkner and Rosen, 1999). It is unclear, however, to what extent any of these linguistic cues may have contributed to the simulated EAS effects observed here.

The effects of both amplitude envelope and voicing (which is a significant component of the amplitude envelope) are not surprising. In general, it is plausible that envelope information in the low-frequency region from either the target or the masker can improve speech intelligibility by providing an indication of when to listen, even at moments when the overall SNR is poor, since at any given moment a relatively favorable SNR is more likely when either the target level is relatively high or the masker level is relatively low (recall that in the present experiment, the low-frequency region did not contain the masker). This glimpsing cue was suggested by Kong and Carlyon (2007) as a possible explanation for the benefit observed in EAS.

Glimpsing might also at least partly explain the effects of F0. That is, F0 might provide an indication of when to listen, much like that provided by the amplitude envelope (Kong and Carlyon, 2007). If F0 and envelope cues both indicate a favorable time to listen, the two cues should be correlated. We evaluated this by comparing the fluctuations in amplitude with those in F0 across ten utterances. Only the voiced segments of each sentence were used, and the F0 track was equated in rms with the envelope track. The lowest r value obtained was 0.29, while the highest was 0.76, and the mean r was 0.52. In general, during the voiced portioned of the sentences we examined, as the amplitude envelope increases, so does F0. This analysis thus suggests that increases in F0 may indicate moments during an utterance in which favorable SNRs are more likely. The fact that they are not more highly correlated is consistent with the finding that F0 and the amplitude envelope appear to provide at least somewhat independent and additive benefit to vocoder alone.

We have clearly demonstrated a benefit in intelligibility due to the presence of F0 in simulated EAS, at least in three of four backgrounds tested. However, it is important to note that as with Kong and Carlyon (2007), the background was never present in the low-frequency region in any of these

conditions. While not ecologically valid, we chose this design because it allowed us a more sensitive measure of the contributions of the cues of interest. However, it is important to determine the effect that the background will have on the improvement observed due to the tone. Experiment 2 addressed this issue by examining the improvement due to the target tone in the low-frequency region with and without the presence of a background tone.

III. EXPERIMENT 2

Experiment 2 was designed to examine the effects of having background information in the low-frequency region. We included conditions in which vocoder stimulation was combined with a tone modulated with F0 and amplitude envelope information from the target talker, a tone carrying these cues from the background talker, or two tones, one carrying the information from the target and one from the background. This allowed us to compare directly the effects of the cues from each source. In addition, conditions in which the vocoder was combined with target and background low-pass speech were included as well, to allow for direct comparison. Thus, we replicated and extended the conditions from Experiment 1 to include conditions containing the low-frequency stimulus representing the background (either low-pass background speech or a tone tracking the F0 and amplitude envelope of the background speech).

A. Method

1. Subjects

Twelve normal-hearing listeners (11 females, 1 male) ranging in age from 26 to 38 years were paid an hourly wage for their services. The criteria for inclusion were identical to those used for Experiment 1, although a different group of listeners was recruited.

2. Stimuli

The processing and hardware were identical to those used in experiment 1. Prior to conducting experiment 1, we had extracted F0 and voicing data from a male (mean F0 =127) production of the CUNY sentences (Boothroyd *et al.*, 1985), as well as from male (mean F0=90) and female (mean F0=184; the target talker in experiment 1) productions of the IEEE sentences for earlier pilot experiments. Because the design of experiment 2 called for F0 data from the background as well as the target, we switched to CUNY target sentences because we were then able to use the two sets of IEEE sentences as background. A consequence of our switch to the CUNY sentence set is that we were able to extend our results from experiment 1 to high-context target materials.

3. Conditions

A target sentence was combined with a background (male or female talker) and processed through a four-channel vocoder (see experiment 1). The output of the vocoder was presented either alone (V) or with a low-frequency cue. In three conditions, the low-frequency cue consisted of 500 Hz low-pass speech, and was either target [or signal (S)] speech alone ($V/S500$),[3] background (B) speech alone ($V/B500$), or both target and background speech ($V/SB500$). Three other conditions combined the vocoder output with either a tone that was modulated in frequency with the target talker's F0 and amplitude with the envelope of the 500 Hz low-pass target speech ($V/S_{\text{F0-env}}$), a tone modulated in the same way using the background's F0 and low-pass envelope ($V/B_{\text{F0-env}}$), or both tones combined ($V/SB_{\text{F0-env}}$).

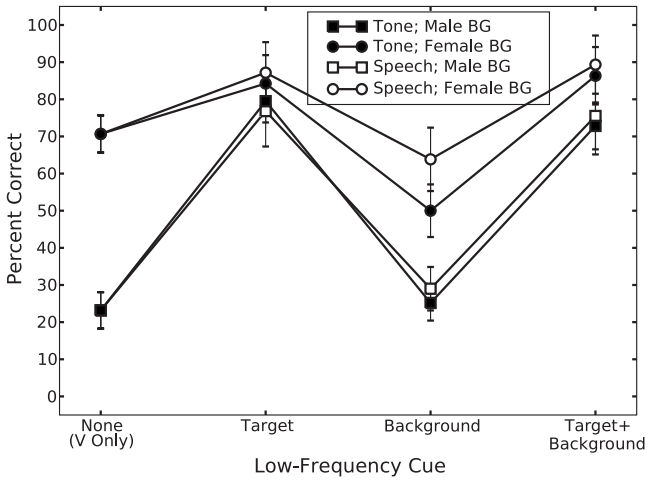

FIG. 2. Mean percent correct scores, ±1 standard error. Squares and circles indicate performance when the background was a male and a female, respectively. Unfilled and filled symbols indicate performance when the low-frequency cue was speech and a modulated tone, respectively. Different low-frequency stimulus conditions are represented along the x axis, and are no low-frequency stimulus (none), target only (target), background only (background), or target and background (target+background). Although the low-frequency stimulus variable, which is depicted along the x axis, is not a continuous variable, the different levels of this variable within each background gender/low-frequency cue combination are connected with lines for clarity.

4. Procedure

The procedure, including the pretest exposure to the vocoder plus target tone condition, was identical to that used in Experiment 1.

B. Results

Figure 2 shows the mean percent correct results. Circles represent performance with a female background and squares represent performance with a male background. Unfilled symbols represent performance when speech was present in the low-frequency region and filled symbols represent performance when a tone or tones were present. The different target/background combinations presented in the low-frequency region are represented along the x axis. There was either no stimulus (vocoder only; none), target only (target), background only (background), or both target and background (target+background). Error bars represent ±1 standard error.

A three-factor repeated-measures analysis of variance was conducted, with background gender (male or female), low-frequency stimulus (none, target, background, or target +background), and low-frequency processing (tone or

speech) conditions as the main effects. There were significant differences within the stimulus and gender variables ($p < 0.001$). Neither the three-way interaction nor any of the two-way interactions were statistically significant (adjusted $p > 0.42$), except the interaction between gender and low-frequency stimulus ($p = 0.001$). This significant interaction is likely due to the differences in performance observed in the vocoder-only conditions (none), which showed the male background to be a more effective masker than the female background. The processing variable was not significant ($p = 0.22$). The lack of a significant difference for processing indicates that a tone conveying both the F0 and the envelope of the target speech provided as much benefit as the low-pass speech itself, whether or not the background was present.

A *post hoc* Tukey analysis was conducted on group means. Within the male background (squares), the presence of the target in the low-frequency region (speech or tone) provided about 55 percentage points of improvement over vocoder alone, regardless of whether the background was present (target+background) or not (target). This improvement was statistically significant (adjusted $p < 0.001$). On the other hand, the low-frequency cue (speech or tone) had no statistically significant effect on intelligibility, regardless of whether or not the target was present (target+background) or not (background) (adjusted $p > 0.99$). Within the female background (circles), there were no statistically significant changes in intelligibility due to either the target or the background, whether the low-frequency stimulus was speech or tones. This effect was likely due to the overall high performance with the female background, which may not have allowed enough room to observe the improvements due to the low-frequency cues.

C. Discussion

The results of Experiment 2 indicate that the EAS benefits of target F0 and amplitude envelope information are not effected by the presence of background F0 and amplitude envelope information. This can be seen by comparing performance in "target" with performance in "target+background." Indeed, the presence of the background in the low-frequency region had no statistically significant effect on performance in any of the comparisons we examined. In other words, performance was equivalent whether the background was present or not. Note, however, the female-background tone condition in which performance declined by about 20 percentage points from vocoder only performance (compare filled circles in "none" and "background").

1. Effects of gender

When the background was a female talker (circles), the amount of improvement due to the low-frequency target stimulus (either speech or tone) was about 15 percentage points. With the male background (squares), which was a much more effective masker in the vocoder-only condition, the improvement in performance when a target stimulus was present in the low-frequency region was about 55 percentage points. At first glance, the difference in masking effectiveness between the female and male backgrounds seems surprising, given that several earlier studies have reported a lack of such an effect for CI patients and normal-hearing subjects listening in simulation (e.g., Qin and Oxenham, 2003; Stickney *et al.*, 2004). This apparent discrepancy may be explained by our use of a sinusoidal vocoder, as most of the studies that report no differences in masking based on gender have used noise-excited vocoders. Consistent with this explanation are recent results from Cullington and Zeng (2008). They used a sinusoidal vocoder and found that speech reception thresholds of a male talker were about 10 dB better in the presence of a female background than in the presence of a male background using a sinusoidal vocoder.

The background gender effect also contrasts with the results from experiment 1, which showed about the same amounts of masking for the male and female backgrounds. This discrepancy is more difficult to explain. F0 separation cannot account for the effects observed; the F0 difference between the target and the male and female backgrounds from experiment 1 are about 102 and 40 Hz, respectively, while for experiment 2 they are 37 and 57 Hz, respectively. Thus, if F0 separation were a critical factor, one would expect a larger background gender effect in experiment 1, but this outcome was not observed. We have conducted a brief pilot experiment[4] to address this issue, which demonstrates that with a sinusoidal vocoder, the ability of one talker to mask another may not be accurately predicted by gender or F0 difference. More work is clearly needed, and we are developing experiments to more fully characterize the relationship between gender, F0 separation, and intelligibility under vocoded conditions.

2. Effects of sentence materials

We have demonstrated that a tone carrying the F0 and amplitude envelope cues from the target can provide significant benefit in simulated EAS, and that the effects of a tone carrying these same cues from a background talker were not significant. In addition to the lack of an effect of low-frequency background cues, experiment 2 provided several other interesting results. First, the amount of benefit due to the tone in experiment 1 was between 24 and 38 percentage points, while in experiment 2 the amount of benefit was as much as 57 percentage points. Second, a tone carrying F0 and the low-frequency amplitude envelope of the target talker provided as much benefit as the low-pass target speech (compare open and filled symbols in the same low-frequency stimulus condition in Fig. 2).

One possible explanation for the difference across experiments 1 and 2 in the amounts of benefit provided by the tone may be related to the amount of context in the target sentences. The CUNY sentences (target materials in experiment 2) are considered to have high context, whereas the IEEE sentences (target materials in experiment 1) are considered to have low context (Duchnowski *et al.*, 2000; Grant and Walden, 1995). It has been shown (e.g., Bell *et al.*, 1992) that the use of high-context sentence materials reduces the dynamic range of intelligibility as compared to low-context sentences, so that a given increase in the amount of information provided yields a correspondingly greater change in per-

cent correct. Another factor may have been the production styles of the target talkers. The IEEE sentence sets were produced with a conversational style, while the CUNY sentence set we used was produced at a slower rate, using a more highly articulated speaking style. The differences in context and speaking styles between the sentence materials may be responsible for the difference in the amount of improvement due to the tone observed between experiments 1 and 2.

IV. GENERAL DISCUSSION

The results of the present experiments demonstrate that F0 can be an important cue under simulated EAS listening. Both F0 and the amplitude envelope contributed significantly to the EAS benefit, and when these two cues were combined, a benefit of as much as 57 percentage points was observed over vocoder only. When sentence context was high, the presence of the tone provided as much benefit as low-pass speech, and this benefit was not adversely affected by the presence of a tone carrying the F0 and amplitude envelope of the low-pass background speech.

The results from Experiment 2, in which the presence of background F0 information provided no benefit to intelligibility, do not support the argument that EAS benefit is due to listeners' improved ability to segregate target and background, as has been proposed (Chang *et al.*, 2006; Qin and Oxenham, 2006; Kong *et al.*, 2005). We would have expected the background to provide some benefit if segregation were the explanation.

On the other hand, intelligibility has been shown to be adversely affected when F0 is flattened (Assman, 1999; Laures and Weismer, 1999) or inverted (Culling *et al.*, 2003). These results have led these investigators to conclude that F0 may help indicate where to listen in an utterance by providing information about where content words are located. This explanation is consistent with the glimpsing account of EAS benefit suggested by Kong and Carlyon (2007), and is a plausible account of our results, although our experiments were not designed specifically to test this.

The benefit provided by a tone carrying F0 and envelope cues observed here in simulations of EAS holds promise for implant listeners, particularly those with significantly elevated thresholds in the low-frequency region. These listeners may not normally benefit from EAS because of their inability to adequately hear a sufficient bandwidth of the speech in the low-frequency region, even with amplification. On the other hand, higher sensation levels (and consequently the potential for some EAS benefit) may be achieved if the low-frequency stimulus were a relatively narrowband tone, modulated in frequency with F0 and in amplitude with the amplitude envelope. It may be possible to construct a processor that extracts F0 in real time, and then applies it (as well as the amplitude envelope) to a tone in the low-frequency acoustic region. Note that this approach is different from previous attempts at exploiting F0 information with electric stimulation. For example, Rosen and Ball (1986) found little benefit from a single-channel implant that conveyed F0. The processor described here would present F0 information in the acoustic region, which would be combined with multiple-channel electrical stimulation in the higher frequency region. This approach is more similar to that used by Faulkner *et al.* (1992), who found that an F0-modulated tone was helpful to profoundly hearing-impaired listeners as an aid to lip reading. The processor described here could greatly expand the population of cochlear implant users who stand to benefit from EAS to include individuals who have very little residual low-frequency hearing. Of course, the efficacy of such a processor would depend on whether the effects observed here in simulation emerge with real implant patients. In that regard, Brown and Bacon (2008) provided promising preliminary data collected on EAS patients that showed that combining electric stimulation with an acoustically presented tone modulated in frequency and amplitude (as done here) can be an effective means of achieving the benefits of EAS.

V. SUMMARY

A tone modulated either in frequency to track the F0 of the target talker, or in amplitude with the amplitude envelope of the target talker provides significant benefit in simulated EAS.

A tone modulated in both frequency and amplitude ($T_{\text{F0-env}}$) generally provides more benefit than either cue alone.

The presence of the tone ($T_{\text{F0-env}}$), under these simulated conditions, resulted in improvements in intelligibility of between 23 and 57 percentage points over vocoder alone. Intelligibility was not affected by the presence of a tone that tracked the F0 and envelope of a background.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant Nos. DC01376 and DC008329 awarded to S.P.B.). The authors gratefully acknowledge Michael Dorman for insightful discussions about this work, Robert Carlyon and an anonymous reviewer for helpful comments on earlier versions of the manuscript, and Bethany Stover for assistance with data collection.

[1]The experimenter who had knowledge of the conditions during testing ran 11 of the participants. The one who did not ran 14 participants. A two-factor analysis of variance was conducted with experimenter as one factor, and processing condition as the other (only the V and $V/T_{\text{F0-env}}$ conditions were included, since any bias would presumably be between these conditions). The interaction term in this analysis was not significant, $p=0.68$, indicating that bias in scoring by the knowledgeable experimenter was not a significant factor in the pattern of results.

[2]Pilot experiments using multi-talker babble background (SNR of +10) indicated that there was a learning effect in the $V/T_{\text{F0-env}}$ condition with a performance asymptote at around 80 sentences (about 30 percentage points of improvement was observed), but no learning was observed in the V condition (400 sentences of exposure yielded about 4 percentage points of improvement). This led to our decision to provide 100 sentences of practice in the $V/T_{\text{F0-env}}$ condition prior to data collection. However, a reviewer expressed concern that the improvements in performance observed in the $V/T_{\text{F0-env}}$ condition, relative to the other tonal conditions, may have been due to this practice. To address this, we compared the learning effects of each of the four tonal conditions (V/T, V/T_{F0}, V/T_{env}, and $V/T_{\text{F0-env}}$) by exposing a separate set of subjects (three per group) to one of the conditions for 100 sentences, then testing them on V, $V/500$,

V/T, V/T_{env}, V/T_{F0}, and $V/T_{F0\text{-}env}$. While we cannot perform inferential statistics on these data due to the small sample size, we can report that overall the pattern of results was similar to that observed in experiment 1 and, moreover, within each learning condition, mean performance was always better in the $V/T_{F0\text{-}env}$ condition than in the condition in which subjects received learning. Subjects who received learning in V/T averaged 25 percent correct when tested in V/T, and 39 percent correct in $V/T_{F0\text{-}env}$. Subjects exposed to V/T_{F0} during learning scored 45 percent correct in V/T_{F0} and 61 percent correct in $V/T_{F0\text{-}env}$. And subjects exposed to V/T_{env} scored 17 percent correct in V/T_{env} and 60 percent correct in $V/T_{F0\text{-}env}$. We therefore conclude that our results were not biased by presenting 100 sentences of the $V/T_{F0\text{-}env}$ condition prior to testing.

[3]In experiment 1 we used T to refer to the tone in the low-frequency region that carried the F0 and envelope information. In experiment 2, T could refer to tones carrying information about the target, the background, or both. Thus, we use S and B to differentiate the target or "signal" from the background.

[4]Under vocoder-only conditions, we combined a female target (mean F0 =211) with either of two female backgrounds (mean F0s of 184 and 240). Thus, the mean F0 "distance" between the target and each background talker was roughly equal (27 and 29 Hz). We found that for the six participants we tested, the background talker whose F0 was lower was a much better masker than the talker with the higher mean F0. While certainly not conclusive, this result demonstrates that neither simple F0 distance nor gender categorization may be enough to predict the amount of masking a particular background might yield with a particular target.

ANSI (**1996**). "ANSI S3.6-1996, Specifications for audiometers," (American National Standards Institute, New York).

Assmann, P. F. (**1999**). "Fundamental frequency and the intelligibility of competing voices," Proceedings of the 14th International Congress of Phonetic Science San Francisco, August.

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

Auditec (**1997**). "Auditory Tests (Revised)," Compact Disc, Auditec, St. Louis.

Bell, T. S., Dirks, D. D., and Trine, T. D. (**1992**). "Frequency-importance functions for words in high- and low-context sentences," J. Speech Hear. Res. **35**, 950–959.

Bird, J., and Darwin, C. J. (**1997**). "Effects of a difference in fundamental frequency in separating two sentences," Paper for the 11th International Conference on Hear., Granthan, UK, August.

Boothroyd, A., Hanin, L., and Hnath, T. (**1985**). "A sentence test of speech perception: reliability, set equivalence, and short term learning (Internal Rep. No. RCI 10)," New York: Speech.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," Ear Hear. **9**, 306–312.

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

Brown, C. A., and Bacon, S. P. (**2008**). "A new approach to electric-acoustic stimulation," J. Acoust. Soc. Am. **123**, 3054.

Chang, J. E., Bai, J. Y., and Zeng, F. (**2006**). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," IEEE Trans. Biomed. Eng. **53**, 2598–2601.

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

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

Cullington, H. E., and Zeng, F. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (**2005**). "Acoustic simulations of combined electric and acoustic hearing (EAS)," Ear Hear. **26**, 371–380.

Duchnowski, P., Lum, D. S., Krause, J. C., Sexton, M. G., Bratakos, M. S., and Braida, L. D. (**2000**). "Development of speechreading supplements based on automatic speech recognition," IEEE Trans. Biomed. Eng. **47**, 487–496.

Faulkner, A., Ball, V., Rosen, S., Moore, B. C., and Fourcin, A. (**1992**). "Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and auditory abilities," J. Acoust. Soc. Am. **91**, 2136–2155.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Gantz, B. J., and Turner, C. (**2004**). "Combining acoustic and electrical speech processing: Iowa/ Nucleus hybrid implant," Acta Oto-Laryngol. **124**, 344–347.

Gantz, B. J., and Turner, C. W. (**2003**). "Combining acoustic and electrical hearing," Laryngoscope **113**, 1726–1730.

Gantz, B. J., Turner, C., Gfeller, K. E., and Lowder, M. W. (**2005**). "Preservation of hearing in cochlear implant surgery: Advantages of combined electrical and acoustical speech processing," Laryngoscope **115**, 796–802.

Grant, K. W., and Walden, B. E. (**1995**). "Predicting auditory-visual speech recognition in hearing-impaired listeners," The 13th International Congress of Phonetic Science, Stockholm, August.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?" J. Acoust. Soc. Am. **109**, 764–774.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Kong, Y., and Carlyon, R. P. (**2007**). "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," J. Acoust. Soc. Am. **121**, 3717–3727.

Kong, Y., Stickney, G. S., and Zeng, F. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

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

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.

Rosen, S., and Ball, V. (**1986**). "Speech perception with the Vienna extracochlear single-channel implant: A comparison of two approaches to speech coding.," Br. J. Audiol. **20**, 61–83.

Spahr, A. J., and Dorman, M. F. (**2004**). "Performance of subjects fit with the advanced bionics CII and nucleus 3G cochlear implant devices," Arch. Otolaryngol. Head Neck Surg. **130**, 624–628.

Spitzer, S. M., Liss, J. M., and Mattys, S. L. (**2007**). "Acoustic cues to lexical segmentation: A study of resynthesized speech," J. Acoust. Soc. Am. **122**, 3678–3687.

Stickney, G. S., Zeng, F., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.

von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Stürzebecher, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss," ORL **61**, 334–340.