Shifting fundamental frequency in simulated electric-acoustic listening

Christopher A. Brown, Nicole M. Scherrer, and Sid P. Bacon

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

(Received 3 August 2009; revised 16 June 2010; accepted 17 June 2010)

Previous experiments have shown significant improvement in speech intelligibility under both simulated [Brown, C. A., and Bacon, S. P. (2009a). *J. Acoust. Soc. Am.* **125**, 1658–1665; Brown, C. A., and Bacon, S. P. (2010). *Hear. Res.* **266**, 52–59] and real [Brown, C. A., and Bacon, S. P. (2009b). *Ear Hear.* **30**, 489–493] electric-acoustic stimulation when the target speech in the low-frequency region was replaced with a tone modulated in frequency to track the changes in the target talker's fundamental frequency (F0), and in amplitude with the amplitude envelope of the target speech. The present study examined the effects in simulation of applying these cues to a tone lower in frequency than the mean F0 of the target talker. Results showed that shifting the frequency of the tonal carrier downward by as much as 75 Hz had no negative impact on the benefit to intelligibility due to the tone, and that even a shift of 100 Hz resulted in a significant benefit over simulated electric-only stimulation when the sensation level of the tone was comparable to that of the tones shifted by lesser amounts. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3463808]

PACS number(s): 43.66.Ts, 43.71.Ky [BCM]

Pages: 1272–1279

I. INTRODUCTION

It is well established that combining either real or simulated electric stimulation with low-frequency acoustic stimulation (electric-acoustic stimulation, or EAS) can improve speech intelligibility significantly (Dorman *et al.*, 2005; Gonzalez and Oliver, 2005; Kong and Carlyon, 2007; Qin and Oxenham, 2006), particularly in background noise. It has also been established that this benefit can be observed whether the low-frequency stimulation occurs in the implanted ear (Gantz and Turner, 2003; Gantz *et al.*, 2005; Turner *et al.*, 2004; von Ilberg *et al.*, 1999) or in the unimplanted ear (Gifford *et al.*, 2007; Kong *et al.*, 2005).

Two recent studies (Cullington and Zeng, 2010; Zhang *et al.*, 2010) have implicated fundamental frequency (F0) in the improvements in speech reception observed in EAS patients. Both studies manipulated the amount of acoustic input to EAS patients by low-pass filtering the speech. One study (Cullington and Zeng, 2010) used a low-pass filter cutoff as low as 150 Hz, and found significant EAS benefit, even though the 150-Hz low-pass speech provided no intelligibility by itself. The second study (Zhang *et al.*, 2010) used cutoffs of 125, 250, 500, and 750 Hz, and also included an unfiltered condition. They found that the 125-Hz condition provided essentially as much benefit as the unfiltered condition (although there was a difference of about 10 percentage points between these two conditions, they were not statistically different). Both studies concluded that F0 was likely the dominant cue, since it is probably the only speech component present at such low frequencies (Hillenbrand *et al.*, 1995).

The results of both of these studies are consistent with other recent suggestions that the benefits of EAS may be due to the presence of F0 cues¹ in the low-frequency acoustic

region (Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006). Indeed, the available evidence seems to point to the importance of F0 cues, and this is consistent with the speech perception literature that shows that F0 cues are important for speech reception in the presence of a competing talker (Assmann, 1999; Assmann and Summerfield, 1990; Bird and Darwin, 1997; Brokx and Nootboom, 1982; Culling and Darwin, 1993). However, most of the available EAS data have been rather indirect.

Recently, the contributions of F0 variation under simulated EAS conditions have been more directly examined, by replacing the target speech in the low-frequency region with either a harmonic complex (Kong and Carlyon, 2007) or a tone (Brown and Bacon, 2009a, 2010) that was modulated in frequency to track the F0 variation of the target talker. Although one study (Kong and Carlyon, 2007) found no additional benefit from the presence of F0 variation, the others (Brown and Bacon, 2009a, 2010) did, and those results have been confirmed with EAS patients (Brown and Bacon, 2009b). In the more recent simulation studies (Brown and Bacon, 2009a, 2010), the frequency of the tone was at the mean F0 of each target sentence, and it was turned on and off with voicing. By adding either F0 variation, the amplitude envelope of the target speech, or both, the independent contribution of each cue, as well as the sum of their contributions, could be observed. Although conditions were run that contained a tone that was unmodulated, the contribution of a static tone located at the mean F0 cannot be determined from these data, because the tone was gated on and off with voicing, and hence the voicing cue was also present.

From the previous studies, it is unclear whether the tone carrying the F0 variation and amplitude envelope cues must be located at the mean F0 of the target talker to be effective. One explanation for how F0 cues aid speech reception in

noise is segregation (Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006). According to this theory, EAS users are able to combine the relatively weak pitch information from the electric region with the strong pitch cue in the acoustic region. This might be taken to suggest that the frequency (representing the pitch) of the tone in the low-frequency region must be equal or nearly equal to the mean F0 of the target talker. One way to test that is to apply the F0 variation and amplitude envelope cues to a tone quite different in frequency from the mean F0 of the target talker. In the current study, the F0 variation is applied to a tone that is *lower in frequency* than the target talker's natural mean F0. If changing the frequency of the tonal carrier has no effect on the EAS benefit, this might be interpreted as evidence against the importance of segregation based on pitch, since the ability to combine the pitch cues from the electric and acoustic regions would presumably depend on the extent to which they are the same.

In addition to determining the contribution of mean F0 to the EAS benefit, the outcome of the current study has potentially important clinical implications. It has been demonstrated that F0 variation is an important cue for EAS (Brown and Bacon, 2009a, 2009b). If the benefits of EAS are shown to be relatively independent of mean F0, then it may be possible to expand the number of patients who could benefit from EAS. For example, a patient with little residual hearing above 125 Hz might not typically benefit from EAS, because the mean F0 of most talkers, particularly women and children, would be inaudible. EAS benefit might be achieved, however, if the tone could be shifted downward in frequency to a region of audibility.

II. EXPERIMENT 1: EFFECT OF SHIFTING THE TONAL CARRIER

A. Methods

1. Participants

Thirteen individuals (one male, 12 females) participated as listeners. They were native speakers of American English and ranged in age from 20 to 50 years. All participants had audiometric thresholds ≤ 20 dB HL at octave and half-octave frequencies from 125–6000 Hz (ANSI, 1996). They were paid an hourly wage for their participation.

2. Stimuli

Target speech consisted of a female talker with a mean F0 of 184 Hz, producing a subset of the IEEE sentences (IEEE, 1969). The competing background was a different female (mean F0=235 Hz) or a male (mean F0=127 Hz), each producing subsets of the AZBIO sentences (Spahr and Dorman, 2004), 4-talker babble (Auditec, 1997), or speech-shaped noise (low-pass filtered at 800 Hz using a 1st-order Butterworth filter). The background was present only in the vocoder region.

A four-channel sinusoidal vocoder was used to simulate electric listening (Brown and Bacon, 2009a). The signal was band-pass 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 (6th-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.

Prior to testing, the target talker's F0 variation was 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. The output of the vocoder was presented either alone, or with a tone in the low-frequency region that carried target F0 variation information. In some cases the tone also carried the amplitude envelope of the 500-Hz low-pass filtered target speech (obtained via half-wave rectification and low-pass filtering at 16 Hz using a 2nd-order Butterworth filter). This carrier tone was always turned on and off with voicing, equated in RMS to the level of the target speech low-pass filtered at 500 Hz. The frequency of the carrier tone was either equal to the mean F0 of the female target (T184 Hz), or shifted downward linearly (via subtraction) in 25-Hz steps to 159, 134, 109 or 84 Hz (T159, T134, T109, and T84, respectively).

3. Procedure

Participants were tested in a double-walled sound booth. Stimuli were delivered using an Echo Gina 3g sound card, Tucker-Davis PA5 attenuators, and Sennheiser HD250 II headphones. The level of the target sentences was adjusted so that when played broadband, their A-weighted slow-peak level was 70 dB SPL. Prior to testing, a signal-to-noise ratio (SNR) was estimated for each subject that would yield approximately 30% correct word recognition in the vocoder-only condition. This was done to avoid floor effects and to ensure sufficient room below ceiling to observe improvement. The SNRs ranged from 6 to 14 dB across subjects. No other level adjustments were made. Participants were instructed to repeat as much of each target sentence as possible. Each sentence contained 5 keywords, with 10 sentences per condition. There were thus 50 keywords per condition. All participants were given a practice session, in which they heard 10 unprocessed sentences, 10 vocoded sentences, and then 100 vocoded sentences combined with a tone modulated in both frequency and amplitude (V/T184) in order to familiarize them with the test materials (Brown and Bacon, 2009a).

4. Listening conditions

The output of the vocoder contained both the target and background at the SNR derived for each subject. The listening conditions were vocoder alone (V), and vocoder plus low-frequency tone (V/T184, V/T159, V/T134, V/T109, and V/T84). The tone was frequency modulated (T_{F0}) or both

FIG. 1. Mean percent correct scores for experiment 1. Panel A represents performance when the low-frequency tone was modulated in frequency with the dynamic changes in the target talker's F0. Panel B represents performance when the tone was also modulated with the amplitude envelope of the low-pass target speech. The frequency of the tonal carrier is along the x axis, and percent correct is along the y axis. Each plot represents a different background.

frequency and amplitude modulated (T_{F0-env}). The target speech (whether vocoder alone or vocoder plus tone) was presented in one of four backgrounds. Because the goal of the experiment was to compare the benefits provided by the various modulated tones, the background was not included in the low-frequency region. This provided a more sensitive measure of the benefits from the cues of interest (Brown and Bacon, 2009a; Kong and Carlyon, 2007). In addition, the presence of a background tone (carrying F0 variation and amplitude envelope cues of the background) in the low-frequency region has been shown to have no effect on the benefit due to the target tone (Brown and Bacon, 2009a). The test conditions were presented in random order for each listener, and no participant heard a target sentence more than once.

B. Results

Figure 1 shows the mean percent correct results. Panel A shows performance when the tone was modulated with the

F0 variation cue only, while panel B shows performance when the tone carried both F0 variation and amplitude envelope cues. Each curve represents performance in a different background, and error bars represent ± 1 standard error. The frequency of the carrier tone is represented along the x axis. Because the pattern of results was similar across backgrounds, all data were collapsed across the four levels of that variable. In the T_{F0} conditions, average performance at the tonal carrier frequencies of 184, 159, 134, 109, and 84 Hz was, respectively, 43, 42, 38, 37, and 33 percent correct. In the T_{F0-env} conditions, the mean performance at each carrier frequency was 44, 44, 43, 42, and 34 percent correct. Mean performance in the vocoder-only condition was 23 percent correct. A two-factor analysis of variance was conducted on carrier frequency (184, 159, 134, 109, and 84 Hz) and processing condition (T_{F0} and T_{F0-env}). This analysis revealed significant main effects for both the carrier frequency variable ($p < .001$), as well as the processing variable ($p < .001$). The interaction between carrier frequency and processing was not statistically significant ($p = 0.57$). The vocoder-only condition was not included in the analysis of variance, because of its non-factorial nature. However, it was included in the post-hoc Tukey test that was conducted on the carrier frequency variable, which included a Holm-Bonferroni correction. This test revealed that performance was statistically equivalent to one another for each carrier frequency condition (adjusted $p > 0.05$), and, with the exception of 84 Hz, statistically different from performance with the vocoder only (adjusted $p < 0.05$). Although speech recognition scores at 84 Hz were not statistically different from those at 184 Hz, they were also not different from those for the vocoder only, so we conclude that the amount of benefit provided by the 84-Hz tone was not as great as at higher carrier frequencies.

C. Discussion

As we reported previously (Brown and Bacon, 2009a), the addition of a low-frequency tone carrying F0 variation or F0 variation and amplitude envelope cues from the target speech provides significant benefit under simulated EAS conditions. We have now demonstrated that the frequency of that carrier tone can be shifted downward by as much as 75 Hz, and still provide as much benefit as the unshifted tone located at the mean F0.

The effects of background on vocoder-only performance in the current experiment are consistent with what we reported previously (Brown and Bacon, 2009a). This is not surprising, given that both the target and background stimuli were the same as used in the previous study. As we mentioned earlier (Brown and Bacon, 2009a), the fact that the speech-shaped noise background was the most effective masker is likely due to the fact that it had more energy in the frequency region encompassed by our vocoder (750–5500 Hz) than did the speech backgrounds.

It has been suggested (Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006) that the benefit observed from EAS may occur because the presence of F0 cues in the low-frequency acoustic region allows the listener to combine

this strong pitch cue with the relatively weak pitch cue from the electric (or vocoder) region to segregate target and masker. Others (Brown and Bacon, 2009a, 2009b, 2010; Kong and Carlyon, 2007) have suggested that F0 variation and amplitude envelope information available in the low-frequency region provide a glimpsing cue that helps the listener know when to listen in the electric (or vocoder) region.

The target talker in the current experiment was a female with a mean F0 of 184 Hz. Considering that equivalent performance was observed whether the tone carrying the acoustic cues was equal to the mean F0 of the target talker or shifted down to 109 Hz, which is well within the normal range of male talkers (Peterson and Barney, 1952), these data seem to provide evidence against a segregation interpretation of EAS based on mean F0 (or voice pitch). If the benefits of EAS are dependent on the listener's ability to combine the strong pitch cue from the low-frequency region with the weak pitch cue from electric stimulation, then there should be less if any benefit when the low-frequency tone is shifted to another frequency. However, even in the presence of conflicting voice pitch information (as was the case in the current experiment when the carrier tone was shifted down in frequency), it is possible that F0 variation is a strong enough cue for segregation, particularly given that other cues, like voicing, would provide a consistent grouping cue to the listener.

On the other hand, these data are not inconsistent with glimpsing as an explanation of EAS, because the glimpsing account relies on the dynamic changes in the target talker's F0 variation (or amplitude envelope cues) to indicate when to listen in the electric region. It has been suggested that in EAS conditions, increases in the amplitude envelope in the low-frequency region provide an indication of when to listen in the electric region, because at those moments, the SNR in the electric region is more likely to be favorable (Kong and Carlyon, 2007). It has also been suggested that F0 variation may play a similar role (Brown and Bacon, 2009a). This hypothesis is consistent with the well-known positive correlation between F0 variation and amplitude envelope. For example, stressed syllables have been shown to be higher in instantaneous F0, as well as in amplitude (Lieberman, 1960).

One noteworthy aspect of the current results is that they are somewhat different from what we have previously reported using similar processing and stimuli (Brown and Bacon, 2009a). In that experiment the improvement over vocoder-only performance due to the combination of voicing, the amplitude envelope, and F0 variation averaged about 30 percentage points across backgrounds. Here, the amount of benefit was between ten and twenty percentage points. We attribute this reduced benefit from the modulated tone to the overall size of the experiment. Experiment 1 contained 48 conditions, which is nearly double the size of Experiment 1 from the previous study (Brown and Bacon, 2009a). In the current experiment, it was clear that subjects were quite fatigued by the end, and thus performance suffered. In addition to the smaller or compressed overall benefit from the modulated tone, the *difference* in benefit between the tone modulated with voicing and F0, and the tone modulated with all three cues was also smaller. Despite this, it is worth reiterat-

FIG. 2. Mean percent correct scores for the follow-up pilot experiment to experiment 1. The frequency of the tonal carrier is along the x axis, and percent correct is along the y axis. The x axis is identical to that of Fig. 1, with the addition of the 84 adjusted data point (84eq), which represents performance when the level of the carrier tone was adjusted to be equal in sensation level (SL) to that of the 184-Hz tone. The background was multi-talker babble.

ing that the difference between the two processing conditions (F0 and F0+env) was statistically significant.

The decline in performance when the carrier tone was shifted down in frequency to 84 Hz (a shift of 100 Hz) is difficult to explain, particularly given that the shift to 109 Hz (a shift of 75 Hz) showed no similar decline. One possible explanation for this finding is that the tone at 84 Hz was simply less audible, given that normal audibility decreases at very low frequencies (ANSI, 1996).

As a first step in addressing the issue of audibility, we ran a pilot experiment with 28 subjects (inclusion criteria identical to experiment 1) using the target and multi-talker background from experiment 1. The T_{F0+env} processing condition from experiment 1 was included, along with the five carrier frequency conditions (the conditions were V/T184, V/T159, V/T134, V/T109, V/T84). In addition, we included a vocoder-only condition, as well as a condition in which the level of the 84-Hz tone was equated in dB sensation level (SL) with that of the 184-Hz tone (V/T84_{eq}). We equated the levels of the tones by taking into account quiet thresholds for 200-ms tones at 84 and 184 Hz. The results are shown in Fig. 2. A one-factor repeated measures analysis of variance was conducted on the processing variable, which revealed an overall significant effect [$F(6,162)=26.01$, $p<0.001$]. A post-hoc Tukey test with a Holm-Bonferroni correction was conducted, which indicated that all of the tone conditions except V/T84 were significantly different from Vocoder Only. The amounts of benefit for the V/T184, V/T159, V/T134, V/T109, V/T84 and V/T84_{eq} conditions were 30, 31, 27, 18, 12, and 22 percentage points, respectively. It is interesting to note that the tone provided about 10 percentage points more benefit than it did in experiment 1. This amount of benefit is similar to what we have observed previously (Brown and Bacon, 2009a). As noted above, the smaller benefit observed here in experiment 1 may have been due to subject fatigue, given the large number of conditions and

hence long experiment duration. Nevertheless, the results of this pilot experiment confirm the results obtained from experiment 1, and extend them to show that performance in simulated EAS may be unaffected by a shift as great as 100 Hz, as long as the SL of the tone is unaffected.

Because this pilot experiment seemed to demonstrate that SL is an important factor when shifting the carrier tone down in frequency, we decided to more fully characterize the relationship between EAS benefit and SL of the modulated tone at various mean F0 values. This was done in Experiment 2.

III. EXPERIMENT II: EFFECT OF SL

A. Methods

1. Participants

A new group of 20 native English speakers (5 males, 15 females) ranging in age from 18 to 37 years were paid for their participation. All participants had audiometric thresholds ≤ 20 dB HL at octave and half-octave frequencies from 125–6000 Hz (ANSI, 1996).

2. Stimuli

Stimuli were generated and processed according to the procedure outlined in experiment 1. The target and background sentences were processed through the four-channel vocoder at a fixed SNR of 10 dB. F0 variation and amplitude envelope cues from the target speech were used to modulate tones at three of the frequencies used in experiment 1 (184, 134, and 84 Hz). At each carrier frequency, the overall level of the tone was manipulated to achieve various SLs (re: quiet thresholds for 200-ms pure tones). Multi-talker babble served as the background for all conditions, and was present in the vocoder region only.

3. Procedure

Quiet thresholds were obtained for 200-ms pure tones at 184, 134 and 84 Hz. These frequencies correspond to the talker's mean F0, and downward shifts of 50 and 100 Hz. The tones were gated with 10-ms raised-cosine ramps. An adaptive two-interval forced-choice procedure was used, with a decision rule that estimates the 70.7% correct point on the psychometric function (Levitt, 1971). Eight reversals were tracked, with step sizes of 5 dB for the first two reversals and 2 dB for the last six, and the signal levels at the last six were averaged to obtain a threshold estimate. Three estimates were obtained for each frequency, and the average of the three was taken as that listener's threshold. The average thresholds across subjects were 17 dB SPL at 184 Hz, 21 dB SPL at 134 Hz, and 28 dB SPL at 84 Hz. Speech intelligibility was measured using a procedure that was identical to that used in experiment 1. The levels for each of the conditions used during testing were checked using a Larson-Davis 800B sound level meter, and the output of the meter was fed to both a Tektronix 2230 scope and a Hewlett Packard 3561A signal analyzer for visual confirmation that the signals were not clipped or otherwise distorted at these levels.

FIG. 3. Mean percent correct scores as a function of the level of the low-frequency tone, in dB SL. Each plot represents performance at a different tonal carrier frequency. The background was multi-talker babble.

4. Listening conditions

Test conditions consisted of V, V/T184, V/T134, and V/T84. The quiet thresholds obtained prior to testing were used as references to present the modulated tone in each processing condition at SLs of 10, 20, 30, 40, 50 and 60 dB for each participant. Test conditions were presented in random order and no participant heard a target sentence more than once.

B. Results

Figure 3 shows the mean percent correct results as a function of the SL of the modulated tone. Error bars represent ± 1 standard error. The different plots represent performance at different carrier frequencies. For carrier frequencies of 184 and 134 Hz, performance increased as SL increased from 10 to 40 dB, and then reached a plateau. The amount of benefit over Vocoder Only was between 25 and 30 percentage points at the highest SLs. When the carrier frequency was 84 Hz, the effect of SL was similar to that for the higher carrier frequencies, although less benefit was observed overall; the maximum benefit was 15 percentage points at 50 dB SL. A two-factor repeated measures analysis of variance was conducted, with carrier frequency (184, 134 and 84 Hz) and SL (10, 20, 30, 40, 50, 60 SL) as the independent variables. A significant main effect was found for carrier frequency ($p < 0.001$) and SL ($p < 0.001$), as well as for the interaction ($p < 0.01$).

We conducted post-hoc Tukey analyses using a Holm-Bonferroni correction. Although the Vocoder-Only condition was not included in the overall analysis, it was included in the post-hoc analyses. These analyses showed that both the 134- and 184-Hz tones provided statistically significant benefit (adjusted $p < 0.05$) over vocoder-only stimulation at all SLs except 10 dB. The amount of benefit observed at these carrier frequencies was statistically equivalent (adjusted $p > 0.80$), which confirms the results obtained in Experiment 1. Performance with the 84-Hz tone was statistically poorer than with the 134-Hz tone (adjusted $p > 0.05$) at all but 10 dB SL. It was also statistically poorer than with the 184-Hz

tone at all but 10 and 20 dB SL. However, the 84-Hz tone did provide statistically significant benefit over vocoder-only stimulation at 30, 40, 50 and 60 dB SL (adjusted $p < 0.05$). In other words, while the 84-Hz tone was not as beneficial as the higher frequency tones, it did provide significant benefit.

C. Discussion

The results of experiment 2 are in general agreement with the main results from experiment 1, which showed that while shifting the tone down by as much as 75 Hz had no effect on performance, the EAS benefit due to the tone was not as great when it was shifted by 100 Hz. In experiment 2, a shift of 50 Hz did not adversely affect performance, although one of 100 Hz did.

Although performance at 84 Hz never reached that observed at 134 or 184 Hz, the improvements over vocoder-only performance at 30, 40, 50 and 60 dB SL were statistically significant at this carrier frequency. That is, even when the tonal carrier was shifted by 100 Hz, significant benefit was observed from the target talker's F0 variation and the amplitude envelope of low-pass target speech, so long as an adequate SL was achieved (at least 30 dB).

Contrary to the results of the pilot experiment reported earlier, the results of experiment 2 indicate that audibility is likely not the only reason for the decline in performance observed in experiment 1 (Fig. 1) when the carrier frequency was shifted downward by 100 Hz. There are several other possible explanations for this decline. One is that performance may be affected by the amount of change in frequency of the tonal carrier, and that a shift of 100 Hz is too large to maintain maximum performance. Another possibility is that the decline in performance may be related to the relative amount of F0 variation, that is, the amount of change around the carrier frequency within an utterance. Our analyses² suggest that variation is normally greater for talkers with higher mean F0 values. When we shifted the tonal carrier down in frequency, however, we did not modify the amount of F0 variation around the mean. In addition, the benefits due to the tone at 84 Hz may have been adversely affected by the lower limits of pitch (Krumbholz, *et al.*, 2000), audibility or both for some of the larger downward excursions even at the highest SLs tested. Informal listening did seem to confirm this possibility, and it is being explored more fully in follow-up experiments.

IV. GENERAL DISCUSSION

The results of the present experiments demonstrate that a tone carrying the F0 variation of the target talker or F0 variation and the amplitude envelope of low-pass target speech can provide a significant benefit in simulated EAS, even when the frequency of the carrier tone is shifted downward by as much as 100 Hz. Experiment 1 showed that downward frequency shifts as large as 75 Hz had no negative impact on performance, while experiment 2 (as well as the pilot experiment associated with experiment 1) showed that the tone can provide significant benefit over vocoder-alone performance even when shifted by 100 Hz as long as an adequate SL is achieved. However, the results from experi-

ment 2 showed that performance when the tone was shifted down by 100 Hz never reached the levels observed for the unshifted tone, or the tone shifted down by 50 Hz. There are several factors that may have contributed to this pattern of results (discussed in Section III C). We are conducting follow-up experiments to examine these possible explanations.

As mentioned in Section II C, a caveat with regard to the data from Experiment 1 is that the pattern of results is somewhat different from both what we have found previously (Brown and Bacon, 2009a), and from the subsequent experiments. That is, the amount of benefit from the modulated tone is somewhat less than what we have observed elsewhere. This difference is likely due to the size and duration of Experiment 1, and the resulting fatigue of the subjects. That is, overall performance in Experiment 1 was depressed because of fatigue, yielding both smaller amounts of benefit due to the various modulated tone conditions than what we have observed elsewhere, as well as smaller differences in performance across processing conditions. This interpretation is supported by the results of both the pilot experiment and Experiment 2, which were both much smaller experiments. In both experiments, the amount of benefit was more similar to what we have observed previously (Brown and Bacon, 2009a).

Recent reports suggest that the percept of F0 cues is stronger with sinusoidal vocoders than with noise vocoders (Stone *et al.*, 2008; Whitmal *et al.*, 2007), and that sinusoidal vocoders may provide similar levels of salience of F0 cues to cochlear implants (CIs; Fu *et al.*, 2005). Because the vocoder implementation used in the current experiments utilized sinusoidal carriers, one would expect that the percept of F0 cues would be relatively good, or nearly as good as can be expected. Those who have proposed that the mechanism for EAS is segregation (Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006), have suggested that the relatively weak representation of F0 cues provided by the CI (or vocoder) are combined with the relatively stronger F0 cues in the low-frequency acoustic region. If segregation is the mechanism for EAS, the sinusoidal vocoder used in the current study should provide a reasonably good opportunity for segregation to occur. However, the results of the current study seem to suggest that segregation based on the pitch in the low-frequency region may not be the mechanism for EAS, since shifting the tone representing mean F0 would seem to eliminate the possibility of grouping based on that cue. However, F0 would still covary between the simulated electric and acoustic regions, and this may be enough to allow grouping of the two sources of information, particularly given that the strong temporal cue of voicing still remains. Despite this possibility, the current study does not provide clear evidence for segregation based on static F0 cues, and does seem to demonstrate that if segregation is the mechanism for EAS, it must be based on dynamic changes in F0 (i.e., F0 variation).

Others have argued that the mechanism is based on glimpsing (Brown and Bacon, 2009a, 2009b, 2010; Kong and Carlyon, 2007; Li and Loizou, 2008). Glimpsing refers to the ability of a listener to take advantage of brief moments

in which the SNR is favorable. Although glimpsing is most often associated with amplitude envelope cues (Kong and Carlyon, 2007; Li and Loizou, 2008), it has been suggested that F0 variation can aid glimpsing as well (Brown and Bacon, 2009a, 2010). According to this theory, F0 variation, along with the amplitude envelope and voicing provide an indication of when to listen in the electric region. For example, voiced segments of speech are generally higher in level than unvoiced segments, and those moments are therefore likely to have a more favorable SNR. Similarly, the moments during which either the amplitude envelope or F0 variation are high may also correspond to more favorable SNRs, and in addition they may indicate stressed syllables, or contextual emphasis. The results from the current experiments seem to support this account, since the benefit of EAS to speech perception is based on F0 variation, and not mean F0. Of course, the current results neither positively confirm nor rule out segregation or glimpsing, and it may be the case that both mechanisms play a role in EAS.

The results of the current study have significant implications for individuals who do not possess enough residual low-frequency hearing to benefit from EAS. One of the reasons an implant patient may not show an EAS benefit is limited hearing in the region in which mean F0 typically occurs. The results of the current experiments suggest that if a patient's residual hearing were restricted to frequencies below the typical range in which mean F0 usually occurs, it may be possible to shift F0 downward in frequency to the region of audibility while preserving amplitude envelope cues. If a wearable real-time processor could be constructed to extract F0 variation and amplitude envelope cues (Faulkner *et al.*, 1992), they could then be applied to a tone that is in an audible frequency region for the patient, thereby providing the advantages of EAS to many individuals who cannot currently achieve such benefit.

V. SUMMARY

1. A tone carrying F0 variation, either alone or in combination with the amplitude envelope of the target talker, provided significant benefit under simulated EAS conditions.
2. Shifting the carrier tone down in frequency by as much as 75 Hz did not affect the benefit provided by the tone.
3. Shifts of 100 Hz also provided significant improvement as long as an adequate sensation level was achieved.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant No. DC008329). The authors gratefully acknowledge Brian Moore and two anonymous reviewers for their helpful comments on earlier drafts of the manuscript.

¹The intended meaning of 'F0' is not always clear; it is often used to describe one or more aspects of voice pitch (Brown and Bacon, 2010). One aspect that is often implied by the use of the term F0 is mean F0, which is the component of voice pitch that is relatively static within a talker, but typically changes across talkers. For example, females typically have a higher mean F0 than males. The term F0 is also often used to describe the dynamic changes that occur in voice pitch across an utterance,

or F0 variation. This is what is removed when speaking in monotone. In the current paper, the terms "mean F0" and "F0 variation" will be used throughout, to avoid confusion. When the more ambiguous use of the term is intended, such as when citing studies that refer to F0 without more specificity, the term 'F0 cues' will be used.

²We computed mean and standard deviation values on the instantaneous changes in F0 variation that occur normally across utterances on all of the speech corpora available to us, including eight different female productions and nine different male productions. The results revealed that more F0 variation was associated with higher mean F0 values. For example, an analysis of a set of IEEE sentences produced by a male talker with a mean F0 of 90 Hz showed that the standard deviation was 16 Hz, which is nearly half the standard deviation of one of the female talkers (Mean F0 = 184 Hz, SD=31 Hz).

- ANSI (1996). American National Standards Specifications for Audiometers (ANSI, New York).
- Assmann, P. F. (1999). "Fundamental frequency and the intelligibility of competing voices," in Proceedings of the 14th International Congress of Phonetic Science, pp. 179–182.
- 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, MO.
- Bird, J., and Darwin, C. J. (1997). "Effects of a difference in fundamental frequency in separating two sentences," in paper for the 11th International Conference on Hearing, Grantham, UK, August.
- Brokx, J., and Nöteboom, S. (1982). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.
- Brown, C. A., and Bacon, S. P. (2009a). "Low-frequency speech cues and simulated electric-acoustic hearing," J. Acoust. Soc. Am. **125**, 1658–1665.
- Brown, C. A., and Bacon, S. P. (2009b). "Achieving electric-acoustic benefit with a modulated tone," Ear Hear. **30**, 489–493.
- Brown, C. A., and Bacon, S. P. (2010). "Fundamental frequency and speech intelligibility in background noise," Hear. Res. **266**, 52–59.
- 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). "The role of timbre in the segregation of simultaneous voices with intersecting F0 contours," Percept. Psychophys. **54**, 303–309.
- Cullington, H. E., and Zeng, F. (2010). "Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear," Ear Hear. **31**, 70–73.
- 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.
- 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.
- Fu, Q., Chinchilla, S., Nogaki, G., and Galvin, J. J. III (2005). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.
- Gantz, B. J., and Turner, C. W. (2003). "Combining acoustic and electrical hearing," Laryngoscope **113**, 1726–1730.
- Gantz, B. J., Turner, C. W., 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.
- Gifford, R. H., Dorman, M. F., McKarns, S. A., and Spahr, A. J. (2007). "Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing," J. Speech Lang. Hear. Res. **50**, 835–843.
- Gonzalez, J., and Oliver, J. C. (2005). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.
- 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.
- 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.
- Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (2000). "The lower limit of pitch as determined by rate discrimination," *J. Acoust. Soc. Am.* **108**, 1170–1180.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Li, N., and Loizou, P. C. (2008). "A glimpsing account for the benefit of simulated combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **123**, 2287–2294.
- Lieberman, P. (1960). "Some acoustic correlates of word stress in American English," *J. Acoust. Soc. Am.* **32**, 451–454.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- 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.
- 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.
- Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (2008). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," *J. Acoust. Soc. Am.* **124**, 2272–2282.
- 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," *J. Otorhinolaryngol. Relat* **61**, 334–340.
- Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (2007). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," *J. Acoust. Soc. Am.* **122**, 2376–2388.
- Zhang, T., Spahr, A. J., and Dorman, M. F. (2010). "Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation," *Ear Hear.* **31**, 63–69.