Published in final edited form as:

Ear Hear. 2016 ; 37(1): e18–e25. doi:10.1097/AUD.0000000000000227.

Shifting fundamental frequency in simulated electric-acoustic listening: Effects of F0 variation

Christopher A. Brown¹, Kate Helms Tillery², Frédéric Apoux³, Nicole M. Doyle², and Sid P. Bacon^{2,4}

Christopher A. Brown: cbrown1@pitt.edu

¹Psychoacoustics Laboratory, Department of Communication Science and Disorders, The University of Pittsburgh, 4033 Forbes Tower, Pittsburgh PA 15260

²Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, PO Box 870102, Tempe AZ 85287-0102

³Speech Psychoacoustics Lab, Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Rd, 110 Pressey Hall, Columbus OH 43210

Abstract

Objective—Shifting the mean fundamental frequency (F0) of target speech down in frequency may be a way to provide the benefits of electric-acoustic stimulation (EAS) to cochlear implant (CI) users whose limited residual hearing precludes a benefit typically, even with amplification. However, previous work showed a decline in the amount of benefit at the greatest downward frequency shifts, and we hypothesized that this might be related to F0 variation. Thus, in the current study we sought to determine the relationship between mean F0, F0 variation, and the benefits of combining electric stimulation from a cochlear implant with low-frequency residual acoustic hearing.

Design—We measured speech intelligibility in normal-hearing listeners using an EAS simulation consisting of a sine vocoder combined either with speech low-pass filtered at 500 Hz, or with a pure tone representing target F0. We used extracted target voice pitch information to modulate the tone, and manipulated both the frequency of the carrier (mean F0), as well as the standard deviation of the voice pitch information (F0 variation).

Results—A decline in EAS benefit was observed at the lowest mean F0 tested, but this decline disappeared when F0 variation was reduced to be proportional to the amount of the shift in frequency (i.e., when F0 was shifted logarithmically instead of linearly).

Conclusion—Lowering mean F0 by shifting the frequency of a pure tone carrying target voice pitch information can provide as much EAS benefit as an unshifted tone, at least in the current simulation of EAS. These results may have implications for cochlear implant users with extremely limited residual acoustic hearing.

Correspondence to: Christopher A. Brown, cbrown1@pitt.edu.

⁴Recently deceased.

INTRODUCTION

The benefits to speech intelligibility in the presence of background noise of combining electric stimulation from a cochlear implant (CI) with low-frequency residual acoustic stimulation (electric-acoustic stimulation, or EAS) are well known (Dorman et al., 2005; Gantz et al., 2005; Gantz and Turner, 2003; Gonzalez and Oliver, 2005; von Ilberg et al., 1999; Kong and Carlyon, 2007; Kong et al., 2005; Turner et al., 2004). We have recently demonstrated that for both EAS patients (Brown and Bacon, 2009b), and normal-hearing listeners in an EAS simulation (Brown and Bacon, 2009a, 2010), the variation in fundamental frequency (F0) of the target talker, as well as the amplitude envelope of the low-pass target speech, are important cues in background noise. These studies demonstrate that the benefit to speech intelligibility from EAS can be observed by replacing the target speech in the low-frequency region with a tone that is modulated in frequency with the target talker's F0 variation, and in amplitude with the amplitude envelope of the low-pass target speech.

Thus, access to the F0 of the target talker, which carries both F0 variation and the amplitude envelope, is useful for successful EAS. However, many CI users do not possess enough residual hearing to show an EAS benefit, and although CIs can deliver amplitude envelope information effectively, there is little F0 variation information present either in real CIs (Chatterjee and Peng, 2008; Wei et al., 2007) or vocoder simulations of CI processing (Chatterjee and Peng, 2008; Stone et al., 2008; Whitmal et al., 2007). As a result, we have begun assessing ways to deliver the cues that are important for the EAS benefit to CI users who do not possess enough residual hearing to access these cues typically, even with amplification. Some of these individuals may have useable residual hearing only in a frequency region below the F0 of most talkers. For example, a CI listener with useable residual hearing up to about 150 Hz listening to a talker whose mean F0 is above that frequency may not show an EAS benefit because the relevant cues in the target speech (F0 variation and low-frequency amplitude envelope) would not be audible. However, if the target speech were replaced with the modulated tone at a frequency below 150 Hz, EAS benefit might be achieved. Other CI users may have an adequate frequency range of low-frequency hearing, but have low-frequency thresholds that are so elevated as to preclude access to F0 variation cues. In either of these cases, it may be possible to provide the benefits of EAS by delivering F0 variation and amplitude envelope cues through alternative means.

We recently began to assess this possibility in simulation for the first group of CI users mentioned: those with a lowered audiometric corner frequency (Brown et al., 2010), using the modulated tone paradigm developed earlier (Brown and Bacon, 2009a). Briefly, a tone is modulated both in frequency with dynamic F0 information from the target speech, and in amplitude with the low-frequency amplitude envelope of the target. EAS benefit was measured while manipulating the frequency of the tone to be either equal to, or less than the mean F0 of the target talker. The goal was to assess the usefulness of the modulated tone at frequencies lower than the natural mean F0 of the target talker. Listeners with normal hearing listened to an EAS simulation, consisting of a sinusoidal vocoder with four logarithmically-spaced channels spanning frequencies from 750 to 5500 Hz, and 500-Hz

low-pass filtered speech. The mean F0 of the target talker was 184 Hz, and the frequency of the carrier tone was set to be 184, 159, 134, 109 or 84 Hz. The extracted F0 variation was applied directly to the tone at each frequency with no further processing, resulting in a linear shift of F0. A statistically significant decline in the amount of EAS benefit was found only at the lowest carrier frequency tested, although even at this frequency, significant EAS benefit was observed. We concluded that it is F0 variation in particular, and not mean F0, that is important for EAS benefit. We also hypothesized that the reduced benefit observed at 84 Hz may be due to reduced audibility at such low frequencies, but found that compensating for the reduced audibility restored only some of the additional benefit that was observed at higher mean F0s (Brown et al., 2010).

Another possible explanation for the reduced EAS benefit may be the relationship between mean F0 and F0 variation. It may be that F0 variation must be reduced to be proportional to the amount of shift in mean F0. That is, F0 should be shifted logarithmically. Brown et al. (2010) used a linear shift, in which F0 variation was fixed, and thus orthogonal with mean F0. The rationale for such a manipulation was threefold. First, it was the simplest implementation, and allowed us to look at the effect of changes to mean F0 while F0 variation was unchanged. In addition, pilot tests at the time showed little decline in benefit using the linear shift, although the lowest mean F0 tested in the study was not used in the pilot. Finally, because the target population for this strategy possess significant hearing loss and would thus likely have broadened auditory filters, and consequently reduced frequency selectivity, we reasoned that the 'expanded' F0 variation that resulted from the linear shift may actually be desirable. There is also indirect evidence for the benefits of expanded F0 from the literature on infant-directed speech (e.g., Fernald, 1985), which shows that it is the exaggerated pitch contour that is the most salient cue (e.g., Fernald and Kuhl, 1987). While this literature does not show a direct benefit to speech acquisition by infants from exaggerated F0 variation (only that infants prefer it), it does suggest that exaggerated F0 may not be deleterious to speech understanding.

Qin and Oxenham (2006) argued that EAS listeners are able to combine the relatively weak envelope pitch cue conveyed by the vocoder (or CI) with the stronger pitch cue in the acoustic region to achieve EAS benefit. Although such fusion would seem to be a necessary precursor for successful EAS, the results reported by Brown et al. (2010) seem to suggest that, at the very least, the mean F0 of the two pitch cues can be different. As discussed in our previous work (Brown and Bacon 2009a, Brown et al 2010), this observation is not inconsistent with the notion that the dynamic F0 changes in the electric and acoustic regions, along with corresponding modulations in the amplitude envelope, may facilitate glimpsing of target speech at moments when signal-to noise-ratios are more favorable. In any case, the limits of lowering F0, and the impact it has on benefit, remain to be seen. The current study was designed to examine several aspects of this issue. First, the issue of how far F0 can be lowered was addressed by a pilot experiment which used a target talker with a relatively high mean F0, to allow for greater downward frequency shifts than those used previously (Brown et al., 2010). Then the question of how F0 lowering should be achieved was explored in Experiments 1 and 2, by independently manipulating mean F0 and F0 variation, and observing their effects on the benefit to speech intelligibility in simulated EAS.

PILOT EXPERIMENT: Effects of Shifting Mean F0

Methods

Participants—Twenty individuals (sixteen females, four males) participated as listeners. They were native speakers of American English and ranged in age from 19 to 34 years. All participants had audiometric thresholds ≤ 20 dB HL at octave and half-octave frequencies from 125 to 6000 Hz (ANSI, 2002). They received either an hourly wage or course credit for their participation.

Stimuli—Target speech consisted of IEEE sentences produced by a female talker with a mean F0 of 238 Hz. Although the CUNY sentence set had been used previously (Brown et al., 2010), we chose this corpus because the relatively high mean F0 of the talker allowed for greater decreases in mean F0, while still remaining safely in the region of commonly occurring voice pitch¹. F0 was extracted from each target sentence using YIN, a well-established pitch algorithm based on auto-correlation (de Cheveigné and Kawahara, 2002), and voice onset/offset information was extracted manually. This procedure is described more fully by Brown and Bacon (2009a).

The competing background was 4-talker babble (Auditec, 1997). Babble, rather than a single talker, was used because our previous work using single-talker backgrounds (Brown and Bacon 2009a) has shown relatively large differences in the amount of masking, depending on the particular talker. As was done previously (Brown and Bacon, 2009a,b; Brown et al., 2010), the background was present only in the vocoder region. We used clean speech in the low-frequency region (below 500 Hz) because we were not interested in testing (nor did we wish to be bound by) the accuracy of the pitch algorithm in noise, but rather how much benefit could be delivered by the modulated tone. Our goal was to examine the change in benefit resulting from changes to target F0, and the addition of a masker would have made it more difficult to do so. It is important to remember this point because it means that the testing conditions, in a sense, were not realistic. In any real-world implementation of the algorithm, noise would be present in the low-frequency region and the results would be different. Thus, our results should not be interpreted as an estimate of how much the addition of a modulated tone would improve speech perception in noise for real-world cochlear implant users.

The four-channel sinusoidal vocoder described by Brown and Bacon (2009a) was used to simulate electric listening. The four bands of the vocoder were logarithmically spaced between 750 and 5500 Hz, and the carriers were sinusoids with frequencies at the arithmetic center of each band. The cutoff frequencies of the envelope filters were 400 Hz, or half the bandwidth, whichever was less. The output of the vocoder was presented either alone, or with a tone presented in the low-frequency region that carried target F0 variation information, and the amplitude envelope of the 500-Hz low-pass target speech (obtained via half-wave rectification and low-pass filtering at 16 Hz with a 2nd-order Butterworth filter). This carrier tone was always turned on and off with voicing. The parameter choices for both

¹The lowest mean F0 observed from among the speech corpora available to us was 86 Hz (see Fig. 4), which was assumed to be a safe estimate of the lower limit of naturally occurring mean F0.

the vocoder and the modulated tone were made to be consistent with our previous EAS work (Brown and Bacon, 2009a, 2010; Brown et al., 2010).

Listening Conditions—The carrier tone described previously was modulated both in frequency with the F0 variation of the target talker, and in amplitude with the amplitude envelope of the 500-Hz low-pass speech. The frequency of the carrier tone was 238 (the mean F0 of the target talker), 188, 138, 113, or 88 Hz (150 Hz below the target talker's mean F0). In all modulated tone conditions except one, the level of the tone was adjusted so that its root-mean-square (RMS) value was equal to that of the 500-Hz low-pass target speech. Given the audibility curve, it is possible that an 88-Hz tone may not be as audible as a higher frequency tone when they are equated in RMS. Thus, one additional 88-Hz modulated-tone condition was added to rule out the issue of audibility, in which the level of the tone was adjusted to be equal in SL to the 238 Hz carrier tone. Additional conditions included one in which vocoder stimulation was presented alone, and one in which vocoder stimulation was combined with 500 Hz low-pass filtered speech (6th-order Butterworth). The presentation order of conditions was randomized for each listener.

Procedure—Participants were tested in a double-walled sound booth. Prior to testing, thresholds in quiet were derived for pure tones at 238 and 88 Hz. These were used to adjust the level of the 88 Hz tone for each listener to be equal in SL to the unshifted tone. All stimuli were delivered using an Echo Layla 3G D/A converter, Tucker-Davis PA4 attenuators, and AKG K271 headphones. For speech testing, participants were instructed to repeat as much of each target sentence as possible. There were 50 keywords per condition. Each target sentence was combined with the 4-talker babble at a signal-to-noise ratio (SNR) of +5 dB, which was shown in pilot testing to yield approximately 30–40% correct vocoder-alone performance. This was done to avoid floor effects and to ensure sufficient room below ceiling to observe improvement. All participants were given a practice session in order to familiarize them with the test materials and processing (Brown and Bacon, 2008). The practice session consisted of 10 unprocessed sentences, 10 vocoded sentences, and then 100 sentences in which the vocoder was combined with the frequency- and amplitude-modulated (FM/AM) tone at a frequency (or mean F0) of 212 Hz. No feedback was provided. As reported previously (Brown and Bacon 2009a), performance with these types of stimulus parameters has been shown to reach asymptote at around 80 sentences. All procedures were approved by the University of Pittsburgh Institutional Review Board.

Results

Results are shown in Fig. 1. The dashed line depicts mean vocoder-only performance, and the solid line shows performance when the vocoder was combined with 500-Hz low-pass filtered target speech (V/500). The plot shows performance in the modulated tone conditions, as a function of carrier tone frequency. Variance indicators represent ± 1 standard error. There is a shallow but steady decline in performance down to a carrier tone frequency of 113 Hz, then performance seems to decline more steeply to 88 Hz. A one-factor analysis of variance showed a significant main effect, $p < .01$. Targeted post-hoc Bonferroni-corrected Tukey pairwise comparisons were made between the vocoder-only condition and each of the modulated tone conditions, and revealed that performance in the vocoder-alone

condition was significantly different from (poorer than) all other conditions except V/T88 and V/T88adj, adjusted $p < .01$. Similarly, all of the modulated tone conditions except the two with carrier tone frequencies of 88 Hz were not statistically different from one another, despite the slight downward trend in performance with decreasing carrier tone frequency (see Fig. 1).

Discussion

Performance over vocoder-alone benefited from the presence of the modulated tone at all carrier frequencies except the lowest tested, 88 Hz, even when the level of the tone was adjusted to be equal in SL to that of the tone at 238 Hz. That is, listeners benefited from the modulated tone even when it was lower in frequency than the mean F0 of the target talker. These results confirm our previous finding that EAS benefit can be observed when mean F0 is lowered (Brown et al., 2010). In addition, adjusting the level of the 88 Hz tone to be equal in SL to the unshifted tone did not result in an increase in the benefit due to the tone.

This prompted us to consider the second hypothesis proposed (Brown et al., 2010), which was that the decline in performance at the lowest carrier tone frequency was due to a disproportionately large amount of F0 variation. In that study, a linear shift in frequency was used as well, the result being that F0 variation is held constant. This may result in too much variation when the frequency of the carrier tone is lowered (See 'Exaggerated F0' in the General Discussion). The alternative would have been a logarithmic shift, which would have resulted in F0 variation being reduced in proportion to the amount of the shift. As described earlier, there were several reasons for using the linear shift, one being that exaggerated F0 variation may prove beneficial for the target population, who have significant hearing loss and thus likely have reduced frequency selectivity.

On the other hand, naturally produced F0 co-modulates with other prosodic speech cues, such as duration and level (Lehiste, 1970). The type of exaggeration resulting from the current processing does not produce co-variations in this way, and thus may be less helpful or even deleterious. This may explain the results of the previous study (Brown et al., 2010) and of the pilot experiment. Subsequent experiments in the current study examined this possibility by lowering the frequency of the carrier tone and reducing the amount of F0 variation. Experiment 1 examined the effects of reducing F0 variation with an unshifted carrier tone. Experiment 2 independently manipulated both F0 variation and mean F0.

EXPERIMENT I: Effects of F0 Variation

Methods

Participants—Twenty individuals (twelve females, eight males) participated as listeners. None had participated in previous experiments. They were native speakers of American English and ranged in age from 18 to 38 years. All participants had audiometric thresholds 20 dB HL at octave and half-octave frequencies from 125 to 6000 Hz (ANSI, 2002). They were paid an hourly wage for their participation.

Stimuli—Target speech consisted of a female talker with a mean F0 of 212 Hz, derived from the CUNY sentences (Boothroyd et al., 1985). The competing background was 4-talker babble (Auditec, 1997). All processing was identical to that used in Experiment 1.

Procedure—The procedure was identical to that used in Experiment 1.

Listening Conditions—All stimuli presented via vocoder stimulation contained both the target and background at an SNR of +5 dB. The listening conditions were vocoder alone (V), vocoder plus 500-Hz low-pass filtered target speech (V/500), and vocoder plus a low-frequency tone (V/T). The tone was either frequency-modulated (TF0) or both frequency- and amplitude-modulated (TF0-env). The amount of frequency modulation (F0 variation), expressed here as multiplication factors in Hz, was manipulated to be 1 (full natural F0 variation), .5 (half the natural F0 variation), .25, .1, 0 (no F0 variation), or -1 (inverted F0 variation). To achieve the reduced F0 variation for each token, a mean was computed of and subtracted from the instantaneous F0 data, to which the desired multiplication factor (.5, .25, etc.) was applied, and finally, the mean F0 was added again.

Results

The results are shown in Fig. 2. Mean percent correct speech intelligibility is plotted as a function of the amount of F0 variation. The dashed and solid lines represent performance in the vocoder-only and V/500 conditions, respectively. The two plots represent performance when the carrier tone was modulated either in frequency only or in both frequency and amplitude. Variance indicators represent ± 1 standard error. A two-factor analysis of variance (F0 variation and processing) was conducted, revealing significant main effects for both factors ($p < .05$), with no significant interaction. Post-hoc Tukey pairwise comparisons, with a Bonferroni correction applied, were also conducted. They revealed that within the TF0-env processing conditions, reversing the F0 variation produced significantly poorer performance than any other condition, while in the TF0 conditions, performance in both the 0 and -1 F0 variation conditions was significantly worse than the others. There were no other significant differences.

Discussion

The post-hoc analyses indicate that, as long as the AM (i.e., amplitude envelope) cue is present, there is no statistically significant decline in performance even when F0 variation is reduced to as little as one-tenth of the natural amount. Taken as a whole, these results suggest a more complex relationship between F0 variation, low-frequency AM, and the benefits of EAS than we have previously suggested (Brown and Bacon, 2009). One way to interpret these results is by considering that in the V/TF0 conditions, the low-frequency AM cue is not removed, but rather replaced by a different AM cue, one that tracks voicing. The decision to always include the voicing cue in the current study is consistent with our previous work that has incorporated the modulated tone paradigm (Brown and Bacon 2009a,b, 2010; Brown et al., 2010), and was made to avoid providing conflicting envelope information to the listener. That is, the tone should not be audible when F0 is not audible. But at particular moments in time when the target talker's F0 is low in level, the absence of the low-frequency amplitude envelope leads to an unduly loud tone, which is another form

of conflicting envelope information. It may be that when the F0 variation cue is reduced, some listeners rely more on the envelope cue or the covariation between the two, which may explain the decline in performance in the V/TF0 conditions.

Nevertheless, a subset of conditions from Experiment 1, when compared to similar conditions from our previous work (Brown and Bacon 2009, 2010), show a relatively consistent pattern of results. Specifically, the presence of both F0 variation and AM cues (V/TF0-env, F0 variation = 1.0; 77 percent correct) provided more benefit than either the F0 cue alone (V/TF0, F0 variation = 1.0; 65 percent correct) or the envelope cue alone (V/TF0-env, F0 variation = 0; 69 percent correct), but not as much as 500 Hz low-pass speech (V/500; 96 percent correct).

Inverting F0 variation had a negative impact on performance. Interestingly, there was a decline in performance between no variation and inverted variation in the TF0-env conditions, but this effect was not present in the TF0 conditions. This is likely because when the envelope was present, the F0 variation conflicted with it, yielding a significant decline. When no envelope was present, there was no conflicting cue, and the presence of inverted F0 variation made no difference.

As mentioned previously (Brown and Bacon 2009a), target sentence context appears to play a large role in overall performance (and the amount of benefit) under these conditions. That effect can be observed in the current study by comparing performance in the Pilot Experiment with that in Experiment 1. In the Pilot Experiment, the IEEE speech corpus (which has low context) was used because the target talker had a particularly high mean F0. In Experiment 1, the CUNY set (which is a high-context corpus) was used. This likely accounts for the relatively large difference in overall performance across the two experiments.

The goal of the current study was to examine whether the decline in performance previously observed when mean F0 was shifted downward (Brown et al., 2010) was due to the variation of F0 being relatively too large. Experiment 1 was designed to establish the effect of reducing F0 variation on the benefit of EAS, so as to provide a basis for interpreting the results of Experiment 2, in which both F0 variation and mean F0 were varied.

EXPERIMENT II: Effects of F0 Variation and Mean F0

Methods

Participants—Twenty eight individuals (sixteen females, twelve males) participated as listeners, none of whom had participated in previous experiments. They were native speakers of American English and ranged in age from 18 to 31 years. All participants had audiometric thresholds ≤ 20 dB HL at octave and half-octave frequencies from 125 to 6000 Hz (ANSI, 2002). They were paid an hourly wage for their participation.

Stimuli—Stimuli and processing were the same as those used in Experiment 1.

Procedure—The procedure was identical to that used in Experiment 1. All procedures were approved by the University of Pittsburgh Institutional Review Board.

Listening Conditions—As with Experiment 1, all stimuli presented via vocoder stimulation contained both the target and background at an SNR of +5 dB. The listening conditions were vocoder alone (V), vocoder plus 500-Hz low-pass target speech (V/500), and vocoder plus a low-frequency tone (V/T). The carrier tone was modulated in both frequency and in amplitude. The frequency of the carrier tone was either the natural mean F0 of the target talker (212 Hz), .5 times the natural mean F0 (106 Hz), or .42 times the natural mean F0 (89 Hz). The latter values were chosen because they are close to the carrier tone frequencies that showed a marked decline in performance previously (Brown et al., 2010). The amount of frequency modulation (F0 variation) was manipulated to be either full natural F0 variation, or reduced to either .5 or .42 times the natural F0 variation. These values were chosen because they yield amounts of variation that are proportional to the amount of shift for each of the mean F0 conditions. It should be noted that when F0 variation is reduced in proportion to a reduction in mean F0, it is equivalent to a logarithmic shift in F0. For convenience in referring to the various processing conditions, a nomenclature will be used whereby, for example, V/T-M89V.5 will refer to the condition in which the mean F0 was 89 Hz and the F0 variation was .5.

Results

The result from Experiment 2 are shown in Fig. 3, in which mean percent correct speech intelligibility is plotted as a function of the amount of F0 variation for each mean F0 condition. As with Experiment 1, vocoder-only and V/500 performance are depicted with dashed and solid lines, respectively. Performance when the mean F0 was 212 Hz (squares) did not decline appreciably as long as some amount of F0 variation was present. This is consistent with the results from Experiment 1. When the mean F0 was 89 Hz (triangles), however, reducing F0 variation caused an improvement in performance of about 10 percentage points. A two-factor within-subjects analysis of variance was conducted, with Mean F0 and F0 variation comprising the two independent variables. The analysis revealed significant main effects for both main factors ($p < .05$), with no significant interaction. Post-hoc Tukey pairwise comparisons, with a Bonferroni correction applied, were then conducted. Conditions included in the analysis were all F0 variation levels when the mean F0 was 89 Hz (specifically, V/T-M89V1, V/T-M89V.5, V/T-M89V.42, and V/T-M89V0), and the V/T-M212V.5 condition, which was added because it represented the highest numerical performance of the M212 conditions. The analysis revealed that performance in the V/T-M89V.42 condition was significantly better than the V/T-M89V0 and V/T-M89V1 conditions, and not statistically different from the V/T-M89V.5 and V/T-M212V.5 conditions.

Discussion

The pattern of results in the three conditions in which F0 variation was not reduced (V1) is similar to that observed previously (Brown et al., 2010), in that as mean F0 was reduced from 212 Hz to 89 Hz, performance declined. We had hypothesized that the decline observed may be due to an unduly large amount of F0 variation at the relatively low mean

F0. In addition to replicating the results from that study, the current study also appears to confirm that hypothesis. That is, when the mean F0 is low, reducing F0 variation results in a further benefit to performance. When F0 variation is reduced to be proportional to the shift in mean F0, the amount of benefit is such that performance is not statistically different from performance at the natural mean F0, at least in the current study. This suggests that when reducing F0, a logarithmic shift (wherein F0 variation is reduced proportionally to the reduction in mean F0) might be the most beneficial way of doing so.

GENERAL DISCUSSION

The relationship between mean F0 and F0 variation

The results of the current study show that when mean F0 is lowered in frequency, optimal speech intelligibility is obtained when F0 variation is reduced proportionally to the F0 shift, at least with the current materials and processing. In addition to behavioral data on the relationship between mean F0 and F0 variation, as is reported here, a quantitative (acoustic) estimate of the relationship between mean F0 and F0 variation also would be informative, and provide convergent validity to our behavioral data. Thus, to determine whether there is, in fact, a positive relationship between mean F0 and F0 variation (as measured by standard deviation), both of these descriptive statistics were computed for all of the speech corpora to which we had access. The only criterion for inclusion was that the speech tokens were naturally-spoken American English sentences produced by a single talker. In all, 33 different talkers, consisting of 17 males and 16 females, were included. The number of tokens per corpus varied from 41 to 866 sentences. For each corpus, the YIN algorithm (de Cheveigné and Kawahara, 2002) was applied to all tokens using a 20-ms step size and no overlap. For each token, a mean and standard deviation for pitch in Hz was computed using all pitch estimates where the probability of the accuracy of the estimate was $\geq .85$, as reported by the algorithm. Those individual token mean and standard deviation values were then averaged across all tokens in a corpus to get a single mean and standard deviation value for each corpus. Fig. 4 plots F0 standard deviation as a function of mean F0. Square symbols represent male talkers, circles represent females. The line is the best-fit regression line with all data (males and females) included. An independent-groups t-test was first computed to confirm that females do have a statistically significantly higher mean F0 than males, $p < .0001$. The grand mean F0 was 118 Hz for the male talkers, and 215 Hz for the females. These results are consistent with previous work on mean F0 and gender (e.g., Fitch and Holbrook, 1970). We then found a statistically significant positive relationship between mean F0 and the amount of F0 variation, $p < .0001$. This result is consistent with the behavioral data from the current study.

Exaggerated F0

It has been known for some time that the exaggerated F0 associated with infant-directed speech is preferred by infants (e.g., Newport, Gleitman and Gleitman, 1977; Fernald and Kuhl, 1987). At first glance, this literature seems to be at odds with the findings of the current study, which suggests that the reason for the decline in intelligibility at lower mean F0s is that the F0 variation is relatively too large. It is likely that the reason 'expanded' F0 caused a decline in performance in the current study is that the other cues associated with

naturally-produced exaggerated F0 (as is observed in infant-directed speech), such as increased mean F0 and decreased speaking rate, etc., were not covaried, which is presumably necessary for benefit.

In any event, the results of the current study demonstrate that linear frequency shifts of F0, as we implemented in an earlier study (Brown et al., 2010), and which would result from frequency transposition, are likely not ideal for lowering mean F0 in an EAS context.

Lowering mean F0 for CI users with limited residual hearing

The results of the current study may have implications for improving speech reception in CI users. We have previously argued that one reason a CI user may not show an EAS benefit is that enough low-frequency hearing may not be available for F0 to be audible. One audiometric configuration for which this might be the case would be when the “corner frequency” is lower than a given target mean F0. In this case, lowering mean F0 into a frequency region of audibility may allow the user to enjoy the benefits of EAS. This possibility is the motivation for our previous work (Brown et al., 2010), as well as the current study. The current results are encouraging, because they suggest that CI patients with the kind of extremely limited residual hearing described earlier may be able to achieve the same amount of EAS benefit as more typical EAS patients. This possibility has yet to be demonstrated with actual CI users, and studies are ongoing.

One limitation of the modulated tone paradigm used in the current study is its reliance on the accuracy of the extracted pitch track. Pitch algorithms are well-known for having trouble in background noise. This was not a problem in the current study, since pitch information was extracted from clean target speech prior to mixing with maskers. This was done because the goal of the study was to examine the change in benefit arising from changes to target F0, and any limits that the pitch algorithm may have had in noise would have made that relationship less clear.

Nevertheless, this is a significant limitation of both the current study, as well as the approach itself, particularly given that the goal of this work is to realize a real-time implementation so that CI users with limited residual hearing can achieve EAS benefits in everyday noisy environments. So the difficulty is clear: it is in noise that the benefits of EAS are greatest, but it is in noise where the type of processing used in the current study will be most difficult to achieve. In light of this, alternative means of frequency lowering are currently being assessed.

The current study lowered mean F0, while either leaving F0 variation intact, or reducing it to be proportional to the F0 shift. One way to think of these two approaches is that the first is equivalent to that which is often termed frequency transposition, while the second would be equivalent to frequency compression. The results of the current study suggest that frequency compression would be the better choice of the two for lowering F0. The possibility of using frequency compression in a hearing aid in conjunction with a cochlear implant has been examined (McDermott and Henshall, 2010), and little benefit was observed. However, that study targeted a different hearing-impaired population, and used an off-the-shelf aid which implemented frequency compression only for frequencies above 1.5 kHz, while frequencies

below that cutoff were not affected. Thus, F0 lowering was not a goal in the above-referenced study. Other studies have examined frequency transposition or compression in hearing aid users and generally found little benefit (e.g., Kuk, et al., 2009; Glista et al., 2009; Simpson et al., 2005), but again, these studies also did not target F0 specifically.

In any event, frequency compression as a means of lowering F0 to augment CI processing appears to hold promise, given the results of the current study. It also has several advantages over the modulated tone paradigm used previously, and in the current study. First, it would be unaffected by the presence of background noise, as the noise would be compressed along with the target signal. Although it should be noted again that the background noise used in the current study was not present in the low-frequency region, so the combined effects of frequency compression and the presence of background noise are yet unknown.

Frequency compression is also computationally inexpensive, and as mentioned, it is already in use in many modern hearing aids. Finally, there exists the possibility that even for CI users with more typical low-frequency acoustic configurations, frequency compression may be advantageous because it may allow additional speech information that is not audible typically, such as F1, to be made audible.

There may be limitations with any approach to lowering F0 for listeners with hearing impairment, however. For example, the frequency response of most commercial hearing aids rolls off somewhere between 250 and 500 Hz. In addition, the two evidence-based prescriptive formulas, DSL 5.0 and NAL-NL2, provide targets above threshold starting at 250 Hz to account, among other things, for issues of loudness. Thus, there is little to no amplification below 250 Hz in commercially-available devices.

These fitting prescriptions and hardware limitations are designed with the goal of making speech as audible and understandable as possible through the hearing aid, and they thus spread the available gain across the speech spectrum. Because F0 is represented in the higher-frequency harmonics of the speech signal, it is not necessary to amplify it directly in typical hearing aids. However, our previous results (Brown and Bacon, 2009a,b; Brown et al., 2010) as well as the results of the current study suggest that in an EAS context, the goals of amplification may be different. That is, they suggest that at least in some cases, it might be beneficial to limit the frequency extent of amplification, even if the patient might be 'amplifiable' at higher frequencies. These studies suggest that the cues carried by F0 are responsible for much of the benefits of EAS, and thus that delivering F0 should be the goal. Given the limited gain available in hearing aids, the difficulty in amplifying low-frequency signals, and the extremely limited residual hearing of the target population, it thus may be desirable to concentrate all of the gain available into the region of F0, to achieve higher sound pressure levels at that frequency region for the given amount of gain, thereby increasing the likelihood that it is audible.

The results of the current study suggest an avenue of research that may result in a means for delivering the cues largely responsible for the benefits of EAS to CI patients who do not receive them typically. Studies are ongoing to assess the feasibility of frequency compression as such an option.

ACKNOWLEDGEMENTS

The authors would like to thank Catherine Palmer for helpful comments on earlier drafts. This work was funded by grant R01-DC008329 from the NIDCD.

REFERENCES

- ANSI. ANSI S12.60-2002, Acoustical performance criteria, design requirements, and guidelines for schools. New York: American National Standards Institute; 2002.
- Auditec. Compact Disc. Auditec, St. Louis; 1997. Auditory Tests (Revised).
- Boothroyd, A.; Hanin, L.; Hnath, T. (Internal No. RCI 10). New York: Speech & Hearing Sciences Research Center, City University of New York; 1985. A sentence test of speech perception: reliability, set equivalence, and short term learning.
- Brown, CA.; Bacon, SP. Learning effects in simulated electric-acoustic hearing. Presented at the 31st Midwinter Meeting of the Association for Research in Otolaryngology; Phoenix, AZ. 2008.
- Brown CA, Bacon SP. Low-frequency speech cues and simulated electric-acoustic hearing. *The Journal of the Acoustical Society of America*. 2009a; 125:1658–1665. [PubMed: 19275323]
- Brown CA, Bacon SP. Achieving electric-acoustic benefit with a modulated tone. *Ear and Hearing*. 2009b; 30:489–493. [PubMed: 19546806]
- Brown CA, Bacon SP. Fundamental frequency and speech intelligibility in background noise. *Hearing Research*. 2010; 266:52–59. [PubMed: 19748564]
- Brown CA, Scherrer NM, Bacon SP. Shifting fundamental frequency in simulated electric-acoustic listening. *J. Acoust. Soc. Am*. 2010; 128:1272–1279. [PubMed: 20815462]
- Chatterjee M, Peng S. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*. 2008; 235:143–156. [PubMed: 18093766]
- de Cheveigné A, Kawahara H. YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am*. 2002; 111:1917–1930. [PubMed: 12002874]
- Dorman MF, Spahr AJ, Loizou PC, Dana CJ, Schmidt JS. Acoustic simulations of combined electric and acoustic hearing (EAS). *Ear and Hearing*. 2005; 26:371–380. [PubMed: 16079632]
- Fitch JL, Holbrook A. Modal vocal fundamental frequency of young adults. *Arch. Otolaryng*. 1970; 92:379–382. [PubMed: 5455579]
- Gantz BJ, Turner C, Gfeller KE, Lowder MW. Preservation of hearing in cochlear implant surgery: advantages of combined electrical and acoustical speech processing. *The Laryngoscope*. 2005; 115:796–802. [PubMed: 15867642]
- Gantz BJ, Turner CW. Combining acoustic and electrical hearing. *The Laryngoscope*. 2003; 113:1726–1730. [PubMed: 14520097]
- Glista D, Scollie S, Bagatto M, et al. Evaluation of nonlinear frequency compression: Clinical outcomes. *Int J Audiol*. 2009; 48:632–644. [PubMed: 19504379]
- Gonzalez J, Oliver JC. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *The Journal of the Acoustical Society of America*. 2005; 118:461–470. [PubMed: 16119365]
- von Ilberg C, Kiefer J, Tillein J, Pfenningdorff T, Hartmann R, Stürzebecher E, Klinke R. Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss. *J. Otorhinolaryngol. Relat. Spec*. 1999; 61:334–40.
- Kong Y, Carlyon RP. Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *The Journal of the Acoustical Society of America*. 2007; 121:3717–3727. [PubMed: 17552722]
- Kong Y, Stickney GS, Zeng F. Speech and melody recognition in binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of America*. 2005; 117:1351–1361. [PubMed: 15807023]
- Kuk F, Keenan D, Korhonen P, et al. Efficacy of linear frequency transposition on consonant identification in quiet and in noise. *J Am Acad Audiol*. 2009; 20:465–479. [PubMed: 19764167]

- Lehiste, I. Suprasegmental features of speech. In: Lass, N., editor. Contemporary Issues in Experimental Phonetics. New York: Academic Press; 1976.
- McDermott H, Henshall K. The use of frequency compression by cochlear implant recipients with postoperative acoustic hearing. *J Am Acad Audiol*. 2010; 21:380–389. [PubMed: 20701835]
- Newport, E.; Gleitman, H.; Gleitman, L. Mother, I'd rather do it myself: some effects and noneffects of maternal speech style. In: Snow, C.; Ferguson, CA., editors. Talking to children: language input and acquisition. Cambridge: Cambridge University Press; 1977.
- Simpson A, Hersbach AA, McDermott HJ. Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *Int J Audiol*. 2005; 44:281–292. [PubMed: 16028791]
- Stone MA, Füllgrabe C, Moore BCJ. Benefit of high-rate envelope cues in vocoder processing: effect of number of channels and spectral region. *The Journal of the Acoustical Society of America*. 2008; 124:2272–2282. [PubMed: 19062865]
- Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA. Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. *The Journal of the Acoustical Society of America*. 2004; 115:1729–1735. [PubMed: 15101651]
- Wei C, Cao K, Jin X, Chen X, Zeng F. Psychophysical performance and Mandarin tone recognition in noise by cochlear implant users. *Ear and Hearing*. 2007; 28:62S–65S. [PubMed: 17496650]
- Whitmal NA, Poissant SF, Freyman RL, Helfer KS. Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *The Journal of the Acoustical Society of America*. 2007; 122:2376–2388. [PubMed: 17902872]

Fig. 1.

Mean speech intelligibility in simulated EAS when the low-frequency target speech was replaced by a tone modulated in both frequency and amplitude with the target talker's F0 information. The parameter is the frequency of the carrier tone, in Hz (which can be thought of as mean F0). The right-most data point in the plot represents performance when the level of the tone at 88 Hz was adjusted to be equal in sensation level to the tone at 238 Hz. The dashed line represents performance with no low-frequency stimulation (V; vocoder only), the solid line represents performance when the vocoder was combined with 500-Hz low-pass filtered target speech (V/500).

Fig 2.

Mean speech intelligibility in simulated EAS when the low-frequency target speech was replaced by a tone modulated in both frequency and amplitude (squares) or in frequency alone (circles) with the target talker's F0 information. The frequency of the carrier tone was equal to the mean F0 of the target talker. The parameter is the amount of F0 variation. The dashed line represents performance with no low-frequency stimulation (V; vocoder only), the solid line represents performance when the vocoder was combined with 500-Hz low-pass filtered target speech (V/500).

Fig. 3. Mean speech intelligibility in simulated EAS when the low-frequency target speech was replaced by a tone modulated in both frequency and amplitude with the target talker's F0 information. The frequency of the carrier tone was either 212 Hz (mean F0 of the target talker; squares), 106 Hz (circles), or 89 Hz (triangles). The amount of F0 variation is varied along the x-axis. The dashed line represents performance with no low-frequency stimulation (V; vocoder only), the solid line represents performance when the vocoder was combined with 500-Hz low-pass filtered target speech (V/500).

Fig. 4. F0 standard deviation as a function of mean F0 for 33 talkers from various sentence corpora typically used in speech intelligibility studies. Males are plotted as squares and females as circles. The solid line is the best-fit linear regression of all data points.