Beneficial acoustic speech cues for cochlear implant users with residual acoustic hearing

Anisa S. Visram,[a] Mahan Azadpour, Karolina Kluk, and Colette M. McKay
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

(Received 15 December 2010; revised 17 February 2012; accepted 27 February 2012)

This study investigated which acoustic cues within the speech signal are responsible for bimodal speech perception benefit. Seven cochlear implant (CI) users with usable residual hearing at low frequencies in the non-implanted ear participated. Sentence tests were performed in near-quiet (some noise on the CI side to reduce scores from ceiling) and in a modulated noise background, with the implant alone and with the addition, in the hearing ear, of one of four types of acoustic signals derived from the same sentences: (1) a complex tone modulated by the fundamental frequency (F0) and amplitude envelope contours; (2) a pure tone modulated by the F0 and amplitude contours; (3) a noise-vocoded signal; (4) unprocessed speech. The modulated tones provided F0 information without spectral shape information, whilst the vocoded signal presented spectral shape information without F0 information. For the group as a whole, only the unprocessed speech condition provided significant benefit over implant-alone scores, in both near-quiet and noise. This suggests that, on average, F0 or spectral cues in isolation provided limited benefit for these subjects in the tested listening conditions, and that the significant benefit observed in the full-signal condition was derived from implantees' use of a combination of these cues.

© *2012 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.3699191]

PACS number(s): 43.66.Ts, 43.71.Ky [RYL] Pages: 4042–4050

I. INTRODUCTION

Cochlear implantees can obtain significant speech perception benefit from combining their implant with residual acoustic hearing in either the implanted or non-implanted ear (Ching *et al.*, 2004; Dorman *et al.*, 2008; Dunn *et al.*, 2005; Gantz and Turner, 2004; Gifford *et al.*, 2007; Gstoettner *et al.*, 2004; Gstoettner *et al.*, 2009; Kiefer *et al.*, 2005; von Ilberg *et al.*, 1999). Support for fundamental frequency (F0) information as a major contributor to bimodal speech benefit has been provided by studies in which the acoustic-side speech was replaced by severely lowpass filtered speech (Cullington and Zeng, 2010; Zhang *et al.*, 2010) or by an F0–modulated tone (Brown and Bacon, 2009b; Carroll *et al.*, 2011). However, Kong and Carlyon (2007) suggested that phonetic cues derived from spectral shape are also present in very low-frequency speech stimuli, and are likely to contribute to bimodal benefit. The goal of this study was to investigate the relative contribution of F0 and spectral cues to bimodal benefit, in conditions approximating quiet (experiment 1) and in a modulated noise (experiment 2). A better understanding of the relative importance of these cues for bimodal speech benefit in different listening conditions may inform hearing aid signal processing strategies designed specifically for bimodal use, by indicating which cues to preserve and potentially strengthen for optimum benefit.

Access to fundamental frequency (F0) information via residual acoustic hearing may be highly informative when listening bimodally (Brown and Bacon, 2009a,b; Carroll *et al.*, 2011; Cullington and Zeng, 2010; Zhang *et al.*, 2010). F0 is poorly coded in implants (e.g., Chatterjee and Peng, 2008) but is, in contrast, likely to be well preserved in low-frequency residual acoustic hearing. Access to F0 provides voicing information, an important cue for consonant identification. According to Stevens' (2002) lexical access model, "acoustic discontinuities" at voicing on/offsets also provide landmark cues that help to segment words and phonemes, providing a structure for lexical recognition. The F0 contour has also been shown to be a useful cue to segmentation (Mattys *et al.*, 2005). Spitzer *et al.* (2009) have shown that, in CI users with residual hearing in the non-implanted ear, removal of the F0 contour of the signal (by presenting the speech with a flattened F0 contour) reduced the ability to segment speech using syllabic stress cues. However, this effect was also evident for bimodal users in the implant-alone condition, showing that it was not reliant on acoustic coding of F0. F0 perception is known to be useful in normal hearing for segregating target speech from a background (Assmann and Summerfield, 1990; Binns and Culling, 2007; Brokx and Nooteboom, 1982). In an amplitude-modulated masker, F0 may help by providing acoustic landmarks and stress cues to give a framework into which glimpsed target information can be integrated in a meaningful way. Access to such landmarks may be particularly important for CI users (Chen and Loizou, 2010), for whom speech information is degraded, especially in noise.

Zhang *et al.* (2010) investigated the benefit of adding lowpass acoustic speech to an implant signal and found that severely lowpass-filtered speech (cut-off 125 Hz) could account for the majority of bimodal benefit. They suggested that this low-passed signal contained only F0 as the major

---

[a] Author to whom correspondence should be addressed. Electronic mail: anisa.visram@manchester.ac.uk

speech cue (for their male speaker) and concluded that perception of F0 plays a major role in bimodal benefit. Cullington and Zeng (2010) also found considerable bimodal benefit due to very low-frequency acoustic information (150 Hz lowpass) and suggested that much of this benefit came from F0 information. It is important to note that even very severely lowpass filtered speech is likely to have some useful phonetic information related to spectral shape. For example, even if the fundamental frequency is the only audible component, it is possible that its amplitude gives some indication regarding the proximity of the first formant, as a higher amplitude F0 would indicate a lower frequency first formant.

In simulations of bimodal hearing in normally hearing listeners, Brown and Bacon (2009a) found that presentation of an acoustic pure tone, modulated to represent the F0 and amplitude envelope contours of the original sentence (AMFM tone), provided speech perception benefit over listening with simulated implant hearing alone. This finding was then replicated in true bimodal users (Brown and Bacon, 2009b) for whom the benefit of adding the AMFM tone was not significantly different to the benefit provided by full acoustic speech presented to the ear with residual hearing. The studies of Brown and Bacon (2009a,b) used relatively low signal-to-noise ratios (SNRs) on the implant side (to limit implant-alone performance to very low levels), and no noise in the acoustic signal, to maximize the effects observable by addition of the F0 cue. Kong and Carlyon (2007) used a method similar to Brown and Bacon to test the effects of F0- and amplitude-modulated complex tones on simulated bimodal speech benefit in normally hearing listeners, with a masker on the simulated-implant side only. Testing at several SNRs, they found benefit of the AMFM complex tone only at the lowest SNR tested (5 dB SNR). Carroll *et al.* (2011) also looked at bimodal speech benefit, using AMFM pure tones to represent the F0 acoustically. They tested CI users on speech perception, with lowpass speech or AMFM tones presented to the ear with residual hearing. For speech perception in quiet (for the portion of listeners who did not show a ceiling effect) the pattern indicated only full speech and not the AMFM tones provided bimodal benefit. For speech perception in the presence of a talker masker (at 10 dB SNR) they found similar benefit whether the lowpass speech + masker was presented acoustically or whether only the target sentence F0 was presented acoustically. In their set-up, only the lowpass speech condition had a masker on the acoustic side, indicating the degree of benefit from acoustic speech in noise was equivalent to the degree of benefit of the AMFM tone presented in quiet. Which particular component of AMFM tones provides benefit is unclear. In simulation studies, Kong and Carlyon (2007) found the benefit from their AMFM tones was accounted for by just the AM component, whereas Carroll *et al.* (2011) found benefit of the FM component only. Brown and Bacon (2009a) found benefits of both AM and FM, as well as benefit of an isolated voicing cue.

As well as accessing F0 information, most bimodal users are able to access some information about the spectral shape of the speech signal, such as low-frequency formant peaks, formant transition slopes and glides associated with dipthongs. This information should aid identification of many vowels and perhaps some consonants such as nasals with low formant frequencies. Kong and Carlyon (2007) found that normally hearing subjects listening to a bimodal simulation (noise vocoder on "implant" side) performed significantly better when listening to lowpass filtered speech on the "acoustic" side than when listening to an amplitude- and F0-modulated complex tone. They inferred that significant phonetic cues from the spectral shape of the 125 Hz lowpass signal were perceived, and integrated with the implant simulation for speech perception benefit. Li and Loizou (2008) showed that low-frequency spectral information is particularly useful for glimpsing target information from a background masker in simulated bimodal hearing.

The amplitude envelope of the low-frequency acoustic signal may also give a cue to voicing, as higher amplitudes at low frequencies tend to indicate voicing. The amplitude envelope is an important cue to manner of consonant articulation, with, for example, plosives indicated by sudden onset or offset of the signal. The amplitude envelope also helps to convey stress patterns which may aid segmentation, particularly in impoverished listening situations (Mattys *et al.*, 2005). It is further possible that the amplitude envelope of the acoustic signal is important for integration of the signals between devices, as amplitude envelope is conveyed quite well in both electric and acoustic hearing. Brown and Bacon (2009a) showed that the amplitude-envelope cue in an acoustic signal can provide independent and additive bimodal benefit to the F0 alone cue, when listening to a bimodal simulation, whereas Carroll *et al.* (2011) found benefit of the FM but not AM component of the tones.

The mechanisms of bimodal benefit are likely to differ depending on whether listening in noise or quiet, as listening in noise requires the listener to segregate the target speech from the background. When considering speech perception benefit in noise, additional hypotheses for bimodal benefit should be considered than when listening in quiet, such as an increased ability to "glimpse" cues in the target speech (Li and Loizou, 2008) and the possibility that cues such as F0 assist in the segregation of target from background, as is well-established in normal hearing listeners (Assmann and Summerfield, 1990). It is possible that listening with noise on the implant side only, as has been the case in some previous studies (Brown and Bacon, 2009a,b; Kong and Carlyon, 2007), exaggerates the usefulness of the acoustic cues being investigated compared to listening situations which are more ecologically valid in terms of the balance of noise between both ears. Presentation of a noise-free modulated tone on the acoustic side is analogous to listening with a hearing aid that has both an F0-extraction algorithm and extremely effective noise reduction algorithm. However, such algorithms generally do not work well at the low SNRs that were used in these studies.

The experimental hypothesis in this study was that both F0 and spectral shape cues contribute to bimodal speech benefit in quiet and in noise. The rationale was that, by processing the speech stimuli to remove particular component cues, the contributions of F0 and spectral cues to bimodal benefit could be isolated. Experiment 1 investigated bimodal speech benefit in near-quiet, specifically focusing on the relative benefit

TABLE I. Demographic data: Age in years; mean audiometric threshold between 125 and 500 Hz; duration of implant use in years and months at start of testing; use of contralateral hearing aid: Y, yes (regular use); N, no use; O, occasional use; SNR is the SNR on the implant side used in experiment 1. Values in brackets are the SNRs used in experiment 1B, for speech-shaped noise and the Dutch vocoded masker, respectively.

ID	Age	Mean PTA/dB HL	Implant use	Hearing Aid use	Processor	Etiology	SNR
S1	68	67	4y8m	Y	Esprit 3G	Familial Progressive	7
S2	78	58	5y8m	O	Harmony	Idiopathic Progressive	15(4,5)
S3	48	58	1y8m	Y	Freedom	Progressive	3
S4	75	58	2y7m	N	Freedom	Familial Progressive	5(1,4)
S5	63	65	4y1m	Y	Freedom	Familial Progressive	12
S6	46	77	1y10m	N	Freedom	Head Injury	7(2,3)
S7	39	20	2y8m	N	Freedom	Idiopathic Progressive	7(−2,1)

provided by different acoustic-side stimuli including: full speech, amplitude- and F0-modulated tones, and vocoded speech. Modulated tones were intended to convey F0 information without spectral shape. Both pure tones (AMFM-pure) and complex tones (AMFM-comp) were used as carriers, thus testing both types of stimulus that have been used to represent F0 cues in previous studies. Vocoded speech was intended to convey spectral shape without F0 information. Experiment 2 compared the relative contribution to speech benefit of these acoustic signals in a modulated noise masker.

II. GENERAL METHODS

A. Subjects

Seven native English-speaking adult post-lingually deafened cochlear implant users participated in experiment 1, six of whom also took part in experiment 2, and four of whom took part in experiment 1B. All participants had residual hearing in the non-implanted ear, with a maximum pure tone threshold at 500 Hz of 90 dB HL. None had aidable residual hearing in the implanted ear. Demographic data are given in Table I and audiometric thresholds from the non-implanted ear are shown in Fig. 1.

B. Target speech

IEEE sentences (IEEE, 1969), produced by a male speaker with a standard Southern English accent, were used as targets. The mean F0 of individual sentences, extracted using the PRAAT speech analysis program (Boersma and Weenink, 2008), ranged between 95 and 128 Hz with a grand mean of 110 Hz.

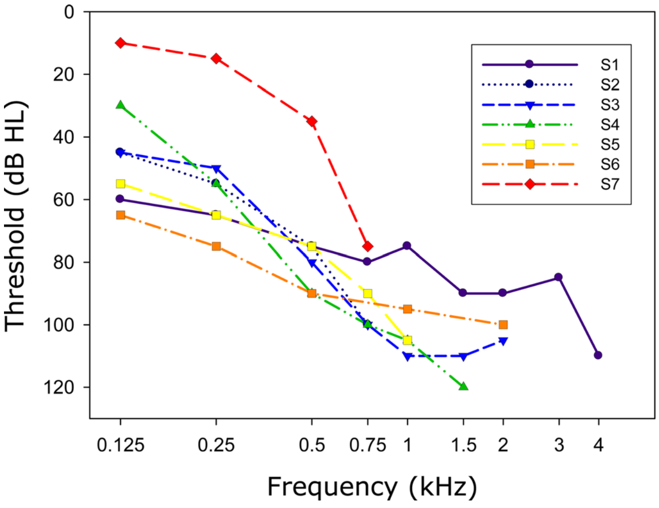

FIG. 1. (Color online) Participants' audiometric thresholds in the non-implanted ear.

C. Acoustic-side stimuli

The non-implanted ear received either no stimulation (implant-alone condition) or acoustic stimuli processed in one of four ways: AMFM-comp, AMFM-pure, vocoded and unprocessed speech. Processing of these stimuli is described below. All acoustic stimuli (including the "unprocessed speech" condition) were first low-pass filtered at 2000 Hz using a 4th-order Butterworth filter before being processed further.

AMFM-comp stimuli represented the information about F0 and amplitude envelope while removing the spectral shape related to phonetic information. First, the pitch contour of the original sentence was extracted using the PRAAT speech analysis program (Boersma and Weenink, 2008). The amplitude envelope of the lowpass portion (800 Hz lowpass, 4th order Butterworth) of the speech signal was then extracted using the Hilbert transform and lowpass filtering at 20 Hz (4th order Butterworth). A 20-harmonic complex was created in MATLAB with an F0 that followed the extracted pitch contour, and amplitude that was modulated by the extracted lowpass amplitude envelope. The signal level was set to zero for the unvoiced speech segments, with 20 ms raised cosine ramps between signal and silence. This signal was passed through a fixed fast-Fourier transform (FFT) filter with the same frequency response as the average of the long-term spectra of the original target sentences. Stimuli were manually checked against the original sentence and corrected for artefacts and missed voicing.

AMFM-pure stimuli were created using the same method as described above but using a pure tone rather than a harmonic complex to represent the F0 and amplitude changes and omitting the speech-shaped filter.

Vocoded stimuli were created in MATLAB using a 14 channel noise-excited vocoder with 150 Hz wide linear bands with a maximum frequency of 2100 Hz. Analysis bands and noise bands were filtered using a 150th-order digital finite impulse response filter (FIR1 in MATLAB). Envelope extraction of the analysis bands was done using the Hilbert transform and lowpass filtering at 20 Hz. The rationale for these parameters was that 150 Hz linear bands would preserve a good degree of spectral shape information, whilst removing harmonic structure (the mean F0 was 110 Hz for our target sentences). The lowpass filtering at 20 Hz was designed to remove any F0 information from the envelope modulations, whilst preserving envelope modulations crucial for intelligibility. The preservation of spectral shape information was checked by extracting vowel segments from the speech stimuli and passing vocoded and non-vocoded vowel extracts through the Auditory Image Model (Bleeck *et al.*, 2004). The overall spectral pattern of these vowel extracts was well-preserved after vocoding (see the Appendix for further details).

In order to adjust the acoustic signals to account for the participant's hearing thresholds, all acoustic stimuli (target sentences across all conditions and all masker stimuli) were initially normalized to an equal average rms value, then passed through an FFT filter based on the frequency response prescribed by the NAL-RP formula for the participant's own audiometric thresholds.

D. Stimulus presentation

The signal in the implanted ear consisted of full (unprocessed) speech. With the exception of s2 presentation to the implant was via direct audio input. Direct input was not possible with s2 as the audio input cable for his device is only advised for use with low-battery-powered devices. Thus, signals were played through a supra-aural headphone placed over the processor's "t-mic" (a microphone adjusted to sit inside the ear canal). Acoustic stimuli were delivered through an EAR 3A insert earphone. Comfortable levels were established prior to testing, using a selection of sentence stimuli that were not used later in the study. A loudness scale was used to identify comfortable levels separately for acoustic and CI stimuli. For each participant the acoustic stimuli, AMFM-comp, vocoded and speech, were all played at the same, comfortable, rms level, and if necessary, extra amplification was provided to AMFM-pure stimuli to achieve a comfortable level. Once monaural comfortable levels were established, if necessary, adjustments were made to both channels to bring the overall level to comfortable and balanced between the two ears when listening bimodally. Such adjustments were necessary for subject s5 only. Stereo files were created in Adobe Audition with the acoustic and implant signals time-locked. These signals were played through a computer connected to an EDIROL UA-25 external sound card.

III. EXPERIMENT 1: SPEECH PERCEPTION IN NEAR-QUIET

A. Procedure

Although the intention of experiment 1 was to investigate sentence perception in quiet, subjects achieved at-or-near-ceiling performance for the sentences in the implant-alone condition making it difficult to assess bimodal benefit. Therefore a low-level speech-shaped noise was added to the implant signal only, to achieve a baseline of approximately 70% keywords correct with implant alone, allowing room for improvement, while approximating speech perception in quiet. To set the noise level, participants listened to IEEE sentences and noise played to the CI ear only, the sentences being fixed to a comfortable level. An adaptive procedure, varying the masker level, was used to find the SNR at which CI-alone scores were roughly 70%. The starting masker level, step size and number of trials varied for each individual depending on their performance and its variability, to provide an estimate of the SNR for 70% correct. This SNR was maintained in the implant signal for all conditions in the test phase of the experiment. The SNRs used for each participant are given in Table I.

For the test phase, 15 new lists (not used to set the SNR level) from the IEEE corpus were chosen at random for each participant (150 keywords over 3 lists for each of 5 conditions). To minimize order effects, stimulus conditions (different acoustic-side stimuli) were interleaved so that on each trial participants heard a different condition to the previous one. Thus there was a loop of 5 conditions continuously repeated. The starting condition within the loop was counterbalanced, as far as possible, between participants. Participants verbally repeated as much of each sentence as they could understand and the experimenter recorded the number of keywords correct.

B. Results

Figure 2 shows individual and mean speech scores for the five experimental conditions. Although the scores are not a continuous variable, lines have been drawn to connect the scores for ease of presentation. A repeated measures analysis of variance (ANOVA) revealed a significant effect of condition ($F(4,6) = 4.687$, $p < 0.001$). *Post hoc* pairwise comparisons (Holm-Sidak method) showed that all conditions differed significantly from CI + speech, with CI + speech giving the better performance ($p < 0.05$) except for the vocoded condition for which the difference only approached significance ($p = 0.058$). No other comparisons were significant, showing that, for this group of subjects, only the full speech condition provided significant benefit on average over listening with the implant alone. There was

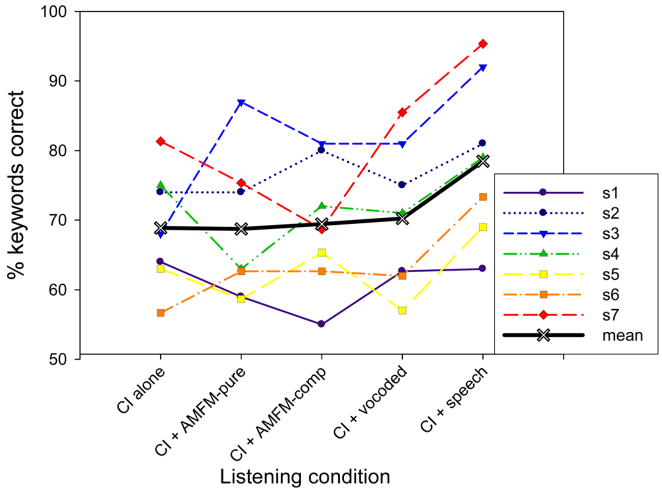

FIG. 2. (Color online) Individual and mean speech scores across conditions in experiment 1.

considerable variability amongst individual participants. For example, one subject (s3) did show significant benefit for all four bimodal conditions over CI alone ($p<0.05$ using binomial analysis, corrected for multiple comparisons) and two showed significant disadvantage of certain conditions (s4 for AMFM-pure and s7 for AMFM-comp, $p<0.05$ using binomial analysis, corrected for multiple comparisons).

C. Experiment 1B: Speech perception with lower SNRs on the implant side

1. Rationale

In contrast with the results of Brown and Bacon (2009b), experiment 1 showed no benefit, on average, of presenting an acoustic tone modulated by the F0 and amplitude envelope of the target signal. The main difference between the studies is that Brown and Bacon used 4-talker babble or single-talker maskers on the implant side at SNRs intended to bring implant-alone scores down to around 20%–30%, whereas only low levels of speech-shaped noise, to bring scores to around 70%, were used in our experiment 1. To further investigate whether the reason for the difference between the current results and those from Brown and Bacon (2009b), were due to the different levels of noise used on the implant side or due to individual subject factors, four participants were able to return to perform speech tests under conditions similar to those used in Brown and Bacon (2009b).

2. Procedure

Four participants (s2, s4, s6, and s7) were tested on IEEE keyword perception with noise on the implant side only, at lower SNRs than those used in experiment 1. Both a speech-shaped noise masker, as used in experiment 1, and a vocoded Dutch speech masker (see experiment 2 for a full description of this masker) were used. For each participant and each masker, the SNR giving 20%–30% correct in the CI alone condition was estimated. This was done with an individually tailored adaptive procedure, as in experiment 1. This SNR for each masker type was maintained in the implant signal for all conditions in the test phase of the experiment, whilst no noise was present on the acoustic side. The SNRs used are shown in Table I. For the test phase, 2 IEEE lists (20 sentences) were presented in each of CI-alone, CI + speech and CI + AMFM-pure conditions, for both types of masker (12 lists in total). One list was completed for each condition and masker before a repeat run in which a second list was completed for each condition and masker (thus the condition changed after 10 sentences, not after each single sentence as in part 1). The order of testing was: speech-shaped noise with CI alone, CI + speech, CI + AMFM-pure, then Dutch masker with CI alone, CI + speech, CI + AMFM-pure. All testing took place in one session.

3. Results

Figure 3 shows the individual and average results, for both noise types. The pattern of results for lower SNRs on the implant side is more similar to that seen in the Brown and Bacon (2009b) study, in which bimodal benefit was derived from modulated tones as well as from the acoustic speech signal. Paired samples t-tests showed a statistically significant benefit of CI + AMFM-pure over CI alone in the low SNR condition for both speech-shaped noise ($t(3) = 19.365$, $p<0.001$) and the Dutch masker ($t(3) = 12.572$, $p<0.001$). Individual trends suggest that all four participants also showed some extra benefit of CI + speech over CI + AMFM-pure for the speech-shaped noise masker, whilst two subjects (s2 and s7) also showed this effect for the Dutch masker, though these effects were not significant in this small group.

IV. EXPERIMENT 2: SPEECH PERCEPTION WITH NOISE IN BOTH EARS

A. Procedure

Previously unheard sentences from the IEEE corpus were used as test material. There were four test conditions comprising all those from experiment one except "AMFM-pure." AMFM-comp was chosen to represent F0 information in this experiment to equate the long-term spectral content between the acoustic-side test conditions and therefore validly compare SNRs between conditions. A series of 21 concatenated sentences spoken by a Dutch male, and passed through a linear 32-channel noise vocoder to remove the fundamental frequency, was used as a masker. The vocoder bandwidth was from 0 to 8000 Hz, with analysis bands and noise bands filtered using a 150th-order digital finite impulse response filter (FIR1 in MATLAB). Envelope extraction of the analysis bands was done using the Hilbert transform and lowpass filtering at 20 Hz. The F0 information was removed to simplify the task for listeners so they were less likely to confuse F0 cues from the target and masker. The masker was added to both implant and acoustic signals at equal SNRs. As none of the participants were Dutch speakers, the masker contained the spectral and amplitude modulation characteristics of speech but was not meaningful.

A one-up one-down adaptive procedure was used to estimate the SNR at which 50% of keywords were correctly identified (SNR50%). The target speech was fixed at a constant comfortable listening level and the masker level was varied. The starting SNR was + 15 dB and was adapted in 5 dB steps for the first two turning points and 2 dB steps for subsequent turning points. A total of 40 trials (4 lists) were used for each run and the SNR was adapted after each sentence (for 0–2 keywords correct the masker level decreased, for 3–5 keywords correct the masker level increased). The SNR50% value was the average SNR of the last 12 turning points. The starting condition was balanced as far as possible between subjects. Two runs of the procedure were performed for each subject, with the repeat set of conditions being tested on a separate day and in the reverse order to that of the first set of conditions. The final SNR50% for each condition was the average of the values obtained from the two runs.

B. Results

Figure 4 shows individual and mean SNR50% values for each condition. As an indication of variability, the mean

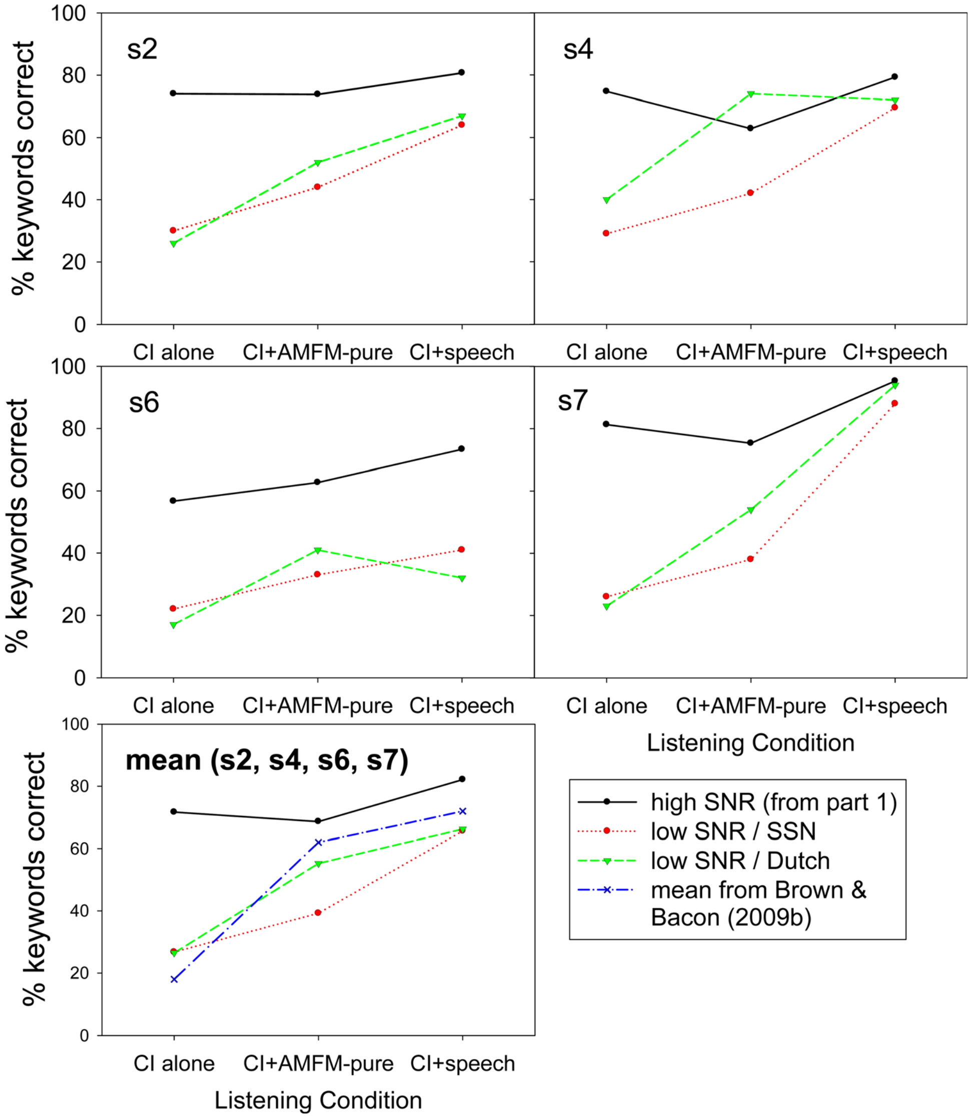

FIG. 3. (Color online) Individual and mean keywords correct for four participants who took part in experiment 1B, with different baseline SNRs on the implant side. "High SNR" results are taken from experiment 1. "Low SNR" results are shown for both noise types: SSN (speech-shaped noise) and Dutch (Dutch vocoded masker). The mean results taken from Brown and Bacon (2009b) have been plotted for comparison.

absolute difference between repeat runs was 1.6 dB. All participants performed best in the CI + speech condition and all but one (s1) performed worst in the CI alone condition. A repeated measures ANOVA revealed a significant main effect of condition ($F(3,4) = 15.14$, $p<0.001$). Pairwise comparisons (Holm-Sidak method) showed that all conditions were significantly worse than CI + speech ($p<0.005$), but that no other comparisons reached significance.

Individual trends suggested that 4 of 6 subjects (s2, s3, s4, and s7) received some bimodal benefit in both CI + vocoded and CI + AMFM conditions (up to 3.2 to 3.8 dB SNR benefit respectively), whilst s6 only gained bimodal benefit in the CI + speech condition, and s1 showed some disadvantage for addition of the AMFM signal.

V. DISCUSSION

In both near quiet (small amount of speech-shaped noise on CI side only, experiment 1) and background noise (modulated masker on both sides, experiment 2) a bimodal benefit was shown when full acoustic speech was presented in addition to the implant signal. However, conditions designed to highlight the specific contribution of two possible sources of bimodal benefit, spectral shape cues and F0 cues, showed no significant bimodal benefit for either type of cue in isolation. For the results with noise presented in both ears, there was a

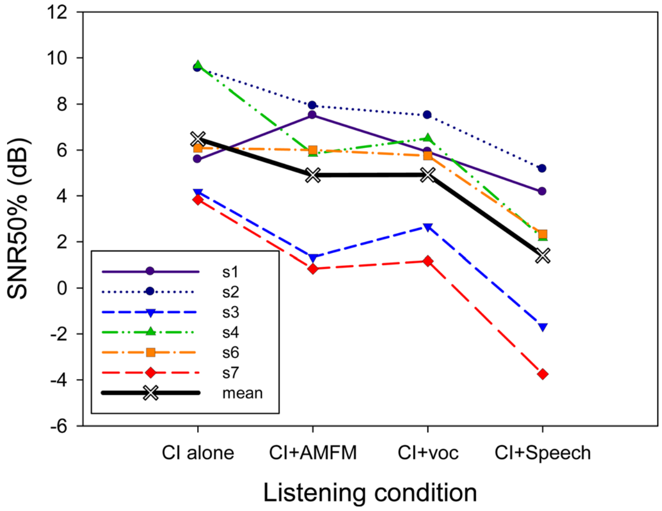

FIG. 4. (Color online) Individual and mean SNR50% values obtained for all participants across conditions in experiment 2. These are the signal-to-noise ratios at which subjects correctly identify 50% of IEEE keywords.

non-significant trend for both partial cues to provide some benefit over CI alone. Lack of significance of these results may in part be due to a lack of statistical power.

Lack of benefit due to AMFM tones is in contrast to the results of Brown and Bacon (2009b) who found significant benefit of AMFM tones, and no significant difference between the benefit given by AMFM tones and full speech. The results of experiment 1B, using noise levels on the CI side comparable to those of Brown and Bacon, did show benefit of AMFM tones, giving some support to the idea that the difference between our results and those of Brown and Bacon (2009b) were due to the differences in the amount of noise used. Figure 3 also suggests the type of noise has an influence on benefit from AMFM tones, with subjects getting more benefit from the AMFM tone in the Dutch (modulated) masker than with speech-shaped noise, though this cannot be supported statistically with the small number of subjects here.

The contrast between the results of experiment 1 and those of Brown and Bacon (2009b) suggests that when the CI signal is impoverished (by addition of noise), the F0/amplitude envelope cue (presented without background noise) is able to provide useful information for bimodal benefit, but when more information is available to the implant, the cues in the AMFM tone become redundant, whereas full speech still provides useful information. From the results of Brown and Bacon (2009b), the extra information available in full speech over AMFM tones seems to be redundant for their subjects and listening conditions, with large amounts of noise on the CI side. It is also plausible that contextual predictability of the speech material had an effect on how well information from the AMFM tones was used. In their simulation study, Brown and Bacon (2009a) showed there was more benefit from the AMFM tone for high predictability sentences than low predictability sentences. In their CI study (Brown and Bacon, 2009b), a mixture of sentence materials with different contextual predictabilities was used whereas only low predictability sentences were used in this study. In experiment 1 the listening condition was changed after each sentence, whereas in experiment 1b, the condition was changed after every 10 sentences. It is possible that this change in procedure could have influenced the results, with subjects having more time to adapt to the AMFM condition in experiment 1b.

Kong and Carlyon's (2007) simulation study showed an AMFM complex tone only gave benefit at the lowest SNR in the simulated CI signal, further suggesting that such signals are most useful when there is poor performance on the implant side. They found this benefit was due to amplitude and/or voicing cues, which they suggested helped to glimpse the target information in the noisy CI signal, whereas the target would always be prominent in quiet listening conditions. The concepts of "glimpsing" and "segmentation" are likely to be closely related when listening in noise, as access to segmental structure is considered an important basis for lexical access (Stevens, 2002), and hence making sense of glimpsed speech segments. The study of Carroll *et al.* (2011) also highlighted how the relative benefit of full speech and AMFM tones for CI users is affected by the noise level used. They only found benefit of AMFM tones (which were presented in quiet) when noise was present on the CI side.

Whilst no evidence of benefit from the isolated F0/amplitude envelope cue was shown in experiments 1 and 2, there was also no evidence of benefit from the isolated spectral cues (i.e., the overall low-frequency spectral shape of the sentences provided by vocoded speech stimuli). The suggestion of Kong and Carlyon (2007) that low-frequency spectral cues were important for simulated bimodal speech benefit was based on their observation of very limited benefit of an isolated F0 and amplitude envelope cue compared to a lowpass speech signal. The current results suggest that benefit from those isolated spectral cues contained in our vocoded stimuli, cannot account fully for the bimodal benefit in the full speech condition. It seems that a combination of spectral and F0 cues is generally necessary for benefit in sentence perception in favorable listening conditions, and in noise.

Zhang *et al.* (2010) showed significant bimodal speech benefit in both quiet and noise of 125 Hz lowpass speech presented acoustically. Although they suggested that this was evidence that F0 cues provided significant benefit, the contrast with the lack of benefit in the F0-only conditions in the current experiment suggests that their lowpass speech stimuli contained more speech information than the AMFM tones used here. It is possible that their listeners were able to use attenuated second harmonic information, and its relation to the first harmonic, to give some cue to the location of the first formant (F1), or even that the amplitude of the first harmonic gave an indication of the F1 peak location, with higher amplitude indicating a lower frequency F1. Thus the amplitude information in their signals may have contained more phonetic information than the AMFM stimuli used in the present study. The study of Kong and Carlyon (2007) also indicated that, when listening to a bimodal simulation, 125 Hz lowpass speech contained more usable speech information than an F0- and amplitude-modulated signal. Another reason for the contrast with the results of Zhang *et al.* (2010) could be that their participants had substantially better residual hearing thresholds than the present ones. They may have therefore been better able to make use of the acoustic F0 cue due to better frequency selectivity.

Perhaps it is not surprising that the CI + speech condition provided the best speech perception, as it combined all available cues in a natural manner. Lack of benefit from the manipulated acoustic signals could suggest that, in near-quiet and noise, the F0 cue and spectral cues may not act independently, but only become beneficial when perceived in combination. It is possible, for example, that F0 is useful to follow the speech signal and to perhaps mark phonemic boundaries, but acoustic spectral information is needed to add content to this structure. When listening in noise, harmonicity in the acoustic signal may help to separate it from the background, with the appropriate spectral shape information then necessary for this extraction to provide useful information. It is further possible that F0 information is better perceived within the context of meaningful spectral information, as formant peaks may allow those harmonics falling within the peak to have sufficient energy to be perceived above a background noise.

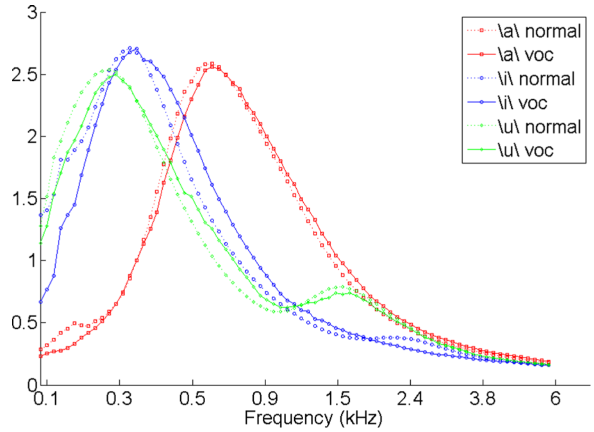

FIG. 5. (Color online) “Stable Auditory Images” derived using the Auditory Image Model for three vowel segments in the normal speech and vocoded acoustic conditions.

There are limits to the idealized concept of functionally separating distinct acoustic cues with these processed stimuli. For example, perceptual interference may have occurred in the AMFM-comp condition if participants typically used spectral information in the low-frequency acoustic signal and were misled by listening to acoustic stimuli with a spectral shape that was inconsistent with that of the target material. Although the vocoded stimuli preserved the overall spectral patterns of the speech stimuli, there will inherently be some loss of fine spectro-temporal detail due to the processing to remove F0. It is possible that the processed conditions may have needed some training for participants to be able to gain benefit from them, and that more practice with these novel stimuli would alter participants’ listening strategies and lead to increased performance. However, a comparison of the first half of trials in experiment 1 with the second half revealed no significant test order effect (the mean difference between first half and second half scores was 0.6 percentage points) and no significant interaction between test order and listening condition. This suggests no systematic learning effects occurred in the course of the study. It is further possible that some benefit could be found from the isolated cues, as was the case for one participant in experiment 1 and was the trend for four participants in experiment 2, but that the sample size was too small to see such effects on average given the individual variability.

There was no clear relationship between bimodal speech benefit and the choice to use amplification in daily life. However, this relationship is difficult to interpret in the present data, given that the group is small and heterogeneous. Some participants who chose not to wear a hearing aid could make use of their natural low-frequency hearing in everyday life without any amplification. This was particularly true for subject 7, who had near-normal thresholds for frequencies up to 250 Hz.

VI. CONCLUSIONS

In both near-quiet and noise, CI users showed significant bimodal benefit from acoustic speech in the non-implanted ear. No significant benefit was observed, across the group of subjects, when acoustic-side speech was replaced with an amplitude- and F0-modulated tone. This lack of benefit of the AMFM tone is in contrast to previous reports, which used noise on the CI side only, and suggests that the benefit of such a tone is not robust when noise is presented to both ears, or in conditions approximating quiet. There was also no significant bimodal benefit from a vocoded acoustic signal, designed to preserve spectral cues whilst removing F0 cues. The results suggest that a combination of F0 and spectral cues is important for bimodal benefit in listening situations with a similar SNR in each device.

ACKNOWLEDGMENTS

This research was financially supported by the University of Manchester. The authors wish to thank the participants for their generous time. Thanks to Astrid van Wieringen for providing the Dutch speech material used, to Martin Vestergaard for help with the using the Auditory Image Model, and to three anonymous reviewers who provided constructive comments on a previous version of the manuscript.

APPENDIX

To investigate the degree of preservation of spectral shape in the vocoded sentences, the Auditory Image Model (AIM) (Bleeck *et al.*, 2004) was used to compare the overall spectral patterns of vowel segments in the vocoded and normal speech conditions. Several vowel segments were taken from parts of the IEEE stimuli, here we present results for three vowel tokens: \a\, \i\, and \u\. The vocoded and normal vowel segments were passed through AIM, with auditory filters broadened by a factor of 5 to simulate loss of frequency selectivity, and without modeling cochlear compression, which we expect to be absent in our participants due to outer hair cell loss. The “stable auditory image” of the vowel segments was generated and plotted in terms of normalized energy across frequencies (normalized to a mean value of 1), which can be seen in Fig. 5. High correlation was found between stable auditory images for the vocoded and normal vowels (for \a\, $r = 0.995$; for \i\, $r = 0.968$; for \u\, $r = 0.989$). First formant peak locations were preserved within 6% of the formant frequency.

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.

Binns, C., and Culling, J. F. (**2007**). “The role of fundamental frequency contours in the perception of speech against interfering speech,” J. Acoust. Soc. Am. **122**, 1765–1776.

Bleeck, S., Ives, T., and Patterson, R. D. (**2004**). “Aim-mat: The auditory image model in MATLAB,” Acta Acust. Acust. **90**, 781–788.

Boersma, P., and Weenink, D. (**2008**). “Praat: Doing phonetics by computer” [Computer program], Version 5.0.27, from http://www.praat.org/ (Last viewed July 4, 2008).

Brokx, J. P., and Nooteboom, S. G. (**1982**). “Intonation and the perceptual separation of simultaneous voices. Jan 1982,” J. Phonetics **10**, 12–36.

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.

Carroll, J., Tiaden, S., and Zeng, F. G. (**2011**). “Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing,” J. Acoust. Soc. Am. **130**, 2054.

Chatterjee, M., and Peng, S. C. (**2008**). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," Hear. Res. **235**, 143–156.

Chen, F., and Loizou, P. C. (**2010**). "Contribution of consonant landmarks to speech recognition in simulated acoustic-electric hearing," Ear Hear. **31**, 259–267.

Ching, T. Y., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Cullington, H. E., and Zeng, F. G. (**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.

Dorman, M. F., Gifford, R. H., Spahr, A. J., and McKarns, S. A. (**2008**). "The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies," Audiol. Neurootol. **13**, 105–112.

Dunn, C. C., Tyler, R. S., and Witt, S. A. (**2005**). "Benefit of wearing a hearing aid on the unimplanted ear in adult users of a cochlear implant," J. Speech Lang. Hear. Res. **48**, 668–680.

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

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.

Gstoettner, W., Helbig, S., Settevendemie, C., Baumann, U., Wagenblast, J., and Arnoldner, C. (**2009**). "A new electrode for residual hearing preservation in cochlear implantation: first clinical results," Acta Otolaryngol. **129**, 372–379.

Gstoettner, W., Kiefer, J., Baumgartner, W. D., Pok, S., Peters, S., and Adunka, O. (**2004**). "Hearing preservation in cochlear implantation for electric acoustic stimulation," Acta Otolaryngol. **124**, 348–352.

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

Kiefer, J., Pok, M., Adunka, O., Sturzebecher, E., Baumgartner, W., Schmidt, M., Tillein, J., Ye, Q., and Gstoettner, W. (**2005**). "Combined electric and acoustic stimulation of the auditory system: Results of a clinical study," Audiol. Neurootol. **10**, 134–144.

Kong, Y. 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.

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.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: a hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.

Spitzer, S. M., Liss, J. M., Spahr, A. J., Dorman, M. F., and Lansford, K. (**2009**). "The use of fundamental frequency for lexical segmentation in listeners with cochlear implants," J. Acoust. Soc. Am. **125**, EL236–EL241.

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

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

Zhang, T., Dorman, M. F., and Spahr, A. J. (**2010**). "Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," Ear Hear. **31**, 63–69.