Review

Fundamental frequency and speech intelligibility in background noise

Christopher A. Brown *, Sid P. Bacon

Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, AZ 85287, United States

ARTICLE INFO

ABSTRACT

Speech reception in noise is an especially difficult problem for listeners with hearing impairment as well as for users of cochlear implants (CIs). One likely cause of this is an inability to 'glimpse' a target talker in a fluctuating background, which has been linked to deficits in temporal fine-structure processing. A fine-structure cue that has the potential to be beneficial for speech reception in noise is fundamental frequency (F0). A challenging problem, however, is delivering the cue to these individuals. The benefits to speech intelligibility of F0 for both listeners with hearing impairment and users of CIs are reviewed, as well as various methods of delivering F0 to these listeners.

Published by Elsevier B.V.

1. The problem of speech in noise

The deleterious effects of background noise on speech intelligibility are well-documented. Although background noise has adverse effects on speech intelligibility for all individuals, those with hearing impairment (HI) are often more severely affected (Bacon et al., 1998; Carhart and Tillman, 1970; Carhart et al., 1969; Festen and Plomp, 1983; Houtgast and Festen, 2008; Plomp, 1978). For example, when speech intelligibility was measured for monosyllables in competing sentences in listeners with normal hearing and listeners with HI, the group with HI exhibited performance that was as much as 20–50 percentage points lower than the group with normal hearing at equivalent signal-to-noise ratios (SNRs) (Carhart and Tillman, 1970). The authors suggested that the effects of sensorineural hearing loss are similar to an increase in masker level relative to the level of the target (i.e., a decrease in SNR) resulting from increased thresholds.

It has since been established, however, that the reduction in intelligibility exhibited by listeners with HI cannot be explained solely by elevated thresholds (Plomp, 1978). This seems to indicate that there are supra-threshold deficits associated with HI. This additional component to the deficits exhibited by listeners with HI was described by a model (Plomp, 1986) that contained two

independent factors: an attenuation factor, and a distortion factor. The attenuation factor represents the threshold shift of a listener with HI. A model based solely on the attenuation factor would show differences in speech intelligibility only at lower noise levels, and a convergence in performance once the level of the noise increased. The distortion factor represents supra-threshold deficits, and specifies that performance will always be worse for listeners with HI, once the level of the noise rises to such a level that it becomes the limiting factor in the audibility of the target speech.

One example of the supra-threshold processing deficits exhibited by listeners with HI can be observed when fluctuating (amplitude-modulated) maskers are used, such as competing speech. It has been known for some time that normal-hearing listeners can take advantage of the momentary favorable SNRs that occur in the temporal valleys of fluctuating maskers to improve speech intelligibility performance while listeners with HI are less able to (Bacon et al., 1998; Duquesnoy, 1983; Festen and Plomp, 1990; Miller and Licklider, 1950; Peters et al., 1998; Plomp, 1994). This ability has been termed "listening in the dips" or "glimpsing".

One report (Festen and Plomp, 1990) that looked at this difference in glimpsing abilities between listeners with normal hearing and listeners with HI measured the SNR required for 50 percent-correct speech recognition (the speech reception threshold, or SRT) for both groups of individuals. The background was varied to be steady-state noise, noise modulated with the amplitude envelope of speech, or a competing single talker. The SRTs for the participants with normal hearing were −4.7 dB in steady-state noise, −8.4 dB for speech-modulated noise, and −11.4 dB for competing speech (see Fig. 1). Thus, the corresponding masking release
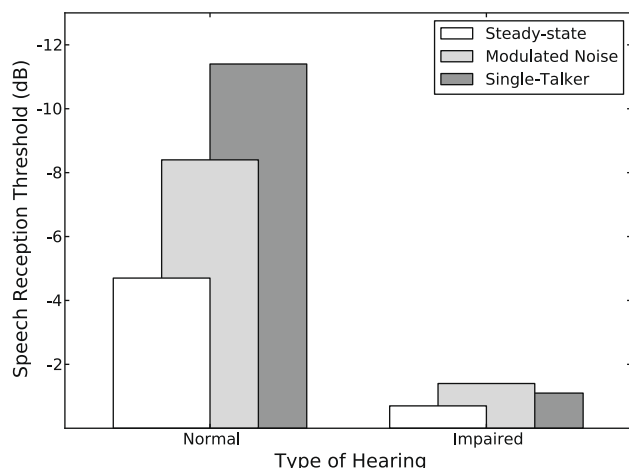
Fig. 1. Data estimated from Fig. 6 in Festen and Plomp (1990). Bars in the left column represent SRT performance by normal-hearing listeners, bars in the right column by listeners with HI. White bars indicate SRTs in steady-state noise, light gray bars indicate SRTs in modulated noise and dark gray indicate SRTs in a single-talker background. For each group of listeners, the difference between the white and gray bars represents the release from masking due to modulation.

(the decline in the effectiveness of the masker) due to an ability to glimpse the target in the valleys of the fluctuating maskers was about 4 dB in speech-modulated noise, and about 7 dB for competing speech. On the other hand, in addition to higher overall SRTs than listeners with normal hearing, listeners with HI showed no release from masking, as evidenced by their nearly equivalent SRTs regardless of the background (Festen and Plomp, 1990).

As adverse as the effects of background noise are for individuals with HI, they are perhaps more adverse for users of cochlear implants[1] (CIs), due to the reduced frequency resolution and lack of fine-structure cues provided by the devices (Cullington and Zeng, 2008; Fu et al., 1998; Nelson and Jin, 2004; Nelson et al., 2003). CIs work by filtering the incoming signal into a number of frequency bands, and extracting the amplitude envelope in each of the bands (Loizou, 1998). Electrodes, positioned at various points along the basilar membrane, emit trains of electrical pulses that are modulated in amplitude with each extracted envelope. Because of this process, fine-structure cues are largely discarded.

In one study that examined the effects of background noise under CI processing (Nelson et al., 2003), speech intelligibility was measured in three groups of participants: users of CIs, normal-hearing listeners listening to broadband stimuli, and normal-hearing listeners who heard stimuli processed with an envelope vocoder, which simulates CI processing.[2] Target speech was heard either in quiet, in a steady noise, or in a gated noise that used a 50% duty cycle. For the gated noise, the modulation rate was varied from 1 to 32 Hz. The results showed that normal-hearing listeners showed releases from masking of as much as 70 percentage points. On the other hand, both normal-hearing listeners in CI simulations and CI patients themselves showed little to no masking release. The authors argue that their results cannot be explained by forward masking, because of the relatively high SNRs used for these groups (+8 and +16). Rather, they suggest that it is the spectrally impoverished nature of

CI processing that prevents a coherent auditory image from forming even in relatively low-level noise, and thus inhibits the listeners' ability to glimpse the target speech (Nelson et al., 2003).

Another study (Stickney et al., 2004) examined speech intelligibility in normal-hearing listeners listening to a CI simulation, and used either speech maskers or noise shaped with the long-term spectra of each respective speech masker. They found no release from masking with speech maskers relative to speech-shaped noise, and in some cases, the speech maskers produced more masking than the noise.

2. The importance of fundamental frequency (F0)

It has been suggested that the lack of masking release in fluctuating maskers (such as competing speech) in both listeners with HI (Festen and Plomp, 1990; Summers and Leek, 1998) and in CI patients (Nelson and Jin, 2004; Qin and Oxenham, 2003; Qin and Oxenham, 2005; Stickney et al., 2004, 2007) is due at least in part to an inability to code F0. This inability likely arises from reduced temporal fine-structure processing (Hopkins and Moore, 2007; Hopkins et al., 2008; Lorenzi et al., 2006, 2009; Moore, 2008; Moore et al., 2006; Stickney et al., 2007). The inability of listeners with HI to make use of temporal fine-structure cues is currently a topic of considerable study, and current CI processing strategies are well-known for discarding all (or nearly all) temporal fine-structure information, and providing only envelope cues to the user.

Although there is ample evidence to support the importance of temporal fine structure and F0 for speech understanding, it is important to note that the redundancy of speech makes temporal fine structure (and F0) neither necessary nor sufficient for good speech reception. It is also the case that not all researchers have argued for the importance of temporal fine structure in the release from masking in modulated backgrounds. For example, when a speech-plus-noise mixture was high-pass filtered at 1500 Hz, the release from masking obtained when the background was modulated was similar to that obtained when the same mixture was low-pass filtered at 1200 Hz (Oxenham and Simonson, 2009). The authors also measured frequency selectivity by obtaining F0 difference limens (DLs) of harmonic complexes that were filtered using the same filter characteristics, and found that the DLs for the high-passed complexes were worse than those for the low-pass complexes by about a factor of 10 (F0 DLs increased from about 1% in the low-pass condition to about 10% in the high-pass condition). Thus, they concluded that F0 was likely not a useful cue in the high-pass conditions, and therefore was not responsible for the release from masking they observed (Oxenham and Simonson, 2009). The results are difficult to interpret, however. Although the DLs are considerably poorer in the high-pass condition than in the low-pass condition, differences in F0 of 10% or more in the high-passed complex were nevertheless detectable. The authors report that the mean F0 of the target talker was 110 Hz, and the standard deviation was 24 Hz, indicating that F0 variation in the target speech commonly exceeded 10% of the mean F0, which was the average DL for the high-pass condition. Thus, this seems to indicate that the F0 variation may have been detectable at least some of the time, even in the high-pass condition where frequency selectivity was greatly reduced.

Despite the conclusions drawn by Oxenham and Simonson (2009), there is evidence from many sources that the availability of the target talker's F0 helps improve intelligibility, particularly in competing speech. F0 has been shown to be important for several linguistic cues, including voicing (Holt et al., 2001; Whalen et al., 1993), lexical boundaries (Spitzer et al., 2007), and manner of articulation (Faulkner and Rosen, 1999). However, it is also an important acoustic cue. In this case, it is useful to consider separately two cues that are often classified under the general term F0: the static mean F0, and the dynamic variations in F0. Mean

---

[1] For clarity, listeners with HI and CI patients are discussed separately throughout, despite the fact that both groups of individuals have HI.

[2] The study of the perception of speech and other acoustic signals by users of CIs has been aided and accelerated by the use of envelope vocoders, which use a processing scheme similar to that of CIs. The main difference between CI processing and vocoder processing is that the electrical pulse train carriers of the CI are replaced with carriers consisting of either bands of noise or pure tones. Because vocoder processing is similar in many respects to CI processing, the vocoder is thought to be a reasonable simulation of CI processing (Dorman et al., 1997; Shannon et al., 1995).

F0 is what is typically thought of when describing the differences in F0 between males and females. F0 variation is the change in F0 that occurs in time across an utterance, and can be thought of as what is removed by speaking in monotone.

2.1. Mean F0

Mean F0 is a well-known cue for speech intelligibility in the presence of competing speech, and it has been shown that intelligibility in listeners with normal hearing increases as mean F0 separation (the difference in mean F0 between target and masker) increases for both vowels (Assmann and Summerfield, 1990; Culling and Darwin, 1993b) and sentences (Assmann, 1999; Bird and Darwin, 1997; Brokx and Nooteboom, 1982; Oxenham and Simonson, 2009; Summers and Leek, 1998). However, there has generally been no benefit to intelligibility from mean F0 separation between target and masker observed in listeners with HI (Summers and Leek, 1998) or in either real or simulated CI listening conditions (Stickney et al., 2004). On the other hand, CI patients have shown some ability to discriminate gender when mean F0 separation was large (Fu et al., 2005). In this study, a closed set of monosyllables was produced by 10 males and 10 females. For each gender, the 5 males with the highest mean F0 were grouped with the 5 females with the lowest mean F0, and the other 10 talkers formed a second group. Thus, there was one group in which the average mean F0 difference between males and females was small (10 Hz), and another group in which the average mean F0 difference between males and females was large (101 Hz). The participants' task was to identify the gender of the talker. Results showed that identification performance by CI patients was about 94 percent correct when the mean F0 difference between males and females was large, and about 68 percent correct when it was small (Fu et al., 2005). It is interesting to note that even in the more difficult condition, performance was still above chance, although the small number of patients tested precluded any statistical analysis.

There have also been a few simulated CI listening studies that have varied mean F0 separation between target and masker by using different speech maskers (Brown and Bacon, 2009b; Cullington and Zeng, 2008). Although both of these studies showed significant differences in speech intelligibility performance with differences in mean F0 separation, they both also used vocoders having sinusoidal carriers, which may provide more pitch information than vocoders using noise-band carriers. In addition, neither of these studies manipulated mean F0 separation per se, but rather used different background talkers. Because the speech materials in these experiments were sentences, this makes interpretation of the effects of mean F0 separation difficult, because more than just mean F0 may vary using this manipulation. For example, speech from different talkers may have been produced with different speaking rates. They may also vary between 'conversational' and 'clear' speaking styles, which has been shown to be a significant factor for speech understanding in both actual CI processing and in simulations of CI processing (Liu et al., 2004). Indeed, both studies conclude that there is not a simple relationship between the amount of masking from speech maskers and mean F0 separation (Brown and Bacon, 2009b; Cullington and Zeng, 2008).

2.2. F0 variation

The effects of F0 variation have been shown to be beneficial for speech reception as well. For example, a relatively early paper demonstrates this effect for listeners with normal hearing (Brokx and Nooteboom, 1982), although it is not a point of emphasis in the paper. A comparison of their Figs. 3 and 5 shows that when the target and masker had different mean F0s ('DIFFERENT PITCH' conditions), the percentage of errors decreased by about 15 per-

centage points on average across SNRs when intonation was normal (with F0 variations) compared to monotonous productions (without F0 variations). Fig. 2 re-plots these data for convenience. The fact that almost no benefit from F0 variation is observed in the 'SAME PITCH' conditions is not surprising, and is likely due to confusion arising from an inability of listeners to discriminate the variations of the target F0 from those of the masker F0. More recently, F0 variation was removed from target speech, which was then presented in a background of either speech-shaped noise or a single talker (Binns and Culling, 2007). While SRTs in speech-shaped noise were equivalent whether the target talker's F0 variation was or was not present, they improved by 2 dB in the single-talker background when the F0 of the target speech varied normally, as compared to when it did not vary at all.

Indirect evidence for a benefit to speech reception from F0 variation may be seen in the results of a few experiments that have noted significant differences in performance depending on the vocoder implementation used. Broadly, vocoders that use sinusoidal carriers have shown more benefit from F0 variation than those that use noise-band carriers (Stone et al., 2008; Whitmal et al., 2007), perhaps because of better F0 representation provided by the sinusoidal carriers (Stone et al., 2008). For example, speech intelligibility in the presence of a competing talker was measured using both tone and noise vocoders (Stone et al., 2008). Although mean F0s are not reported, the target and masker were both males who had considerable overlap in the reported ranges of F0. In addition to vocoder carrier type, the low-pass cutoff frequency of the envelope extraction filters was also varied to be either 45 or 180 Hz. This manipulation provided a signal that either contained (180 Hz) or did not contain (45 Hz) F0 information from periodicity cues in the envelope. While the differences in performance due to envelope cutoff frequency were relatively small (around 10 percentage points) when the vocoder used noise-band carriers, the same differences were as much as 20–25 percentage points with the sinusoidal vocoder (see Fig. 3). These results seem to indicate that the F0 cue was better preserved when the vocoder carriers were tones. In addition to cutoff frequency, the slope of the envelope extraction filter has also been shown to affect intelligibility (Healy and Steinbach, 2007) when sinusoidal carriers are used. In this study, speech reception was measured using a 3-channel sinusoidal vocoder and a male target talker. The cutoff frequency of the envelope extraction filter was either 16 or 100 Hz. When the slope of the envelope extraction filter was decreased from 96 to 12 dB/octave, performance improved significantly (Healy and Steinbach, 2007).
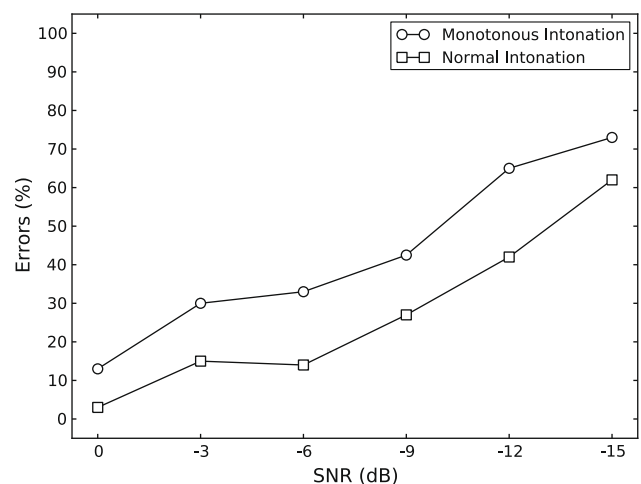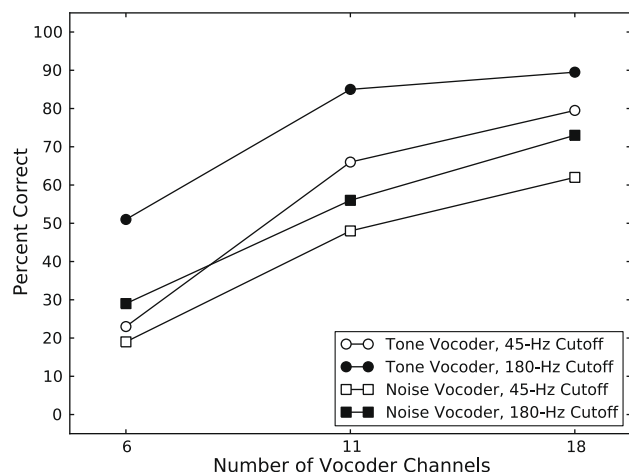
Fig. 2. Data estimated from Figs. 3 and 5 in Brokx and Nooteboom (1982). Percent errors as a function of SNR when intonation was either monotonous (circles) or normal (squares). The percent errors decreased by an average of about 15 percentage points across SNRs with the addition of normal intonation.

Fig. 3. Data estimated from Fig. 3 in Stone et al. (2008). Mean percent correct intelligibility as a function of number of vocoder channels. Plots with square symbols represent performance when the vocoder used noise-band carriers; plots with circles represent sinusoidal carriers. Open symbols represent performance when the cutoff frequency of the envelope extraction filter was 45 Hz, whereas filled symbols represent performance when the cutoff was 180 Hz.

Another recent study (Chatterjee and Peng, 2008) compared the abilities to discriminate intonation patterns in speech of both listeners with normal hearing listening to a vocoder, and users of CIs. English-speaking listeners heard the word 'popcorn,' which was resynthesized to contain either a rising intonation pattern, which in English would indicate a question, or a falling intonation pattern, which would indicate a statement. Their task was to indicate whether the stimulus was a statement or question. Results clearly demonstrated that listeners could discriminate the different intonation patterns. These researchers (Chatterjee and Peng, 2008) found a similar pattern of results for listeners with CIs as well, despite considerable variability across users in their ability to discriminate the patterns of intonation.

One report (Wei et al., 2004) that examined the ability to use F0 variation in both listeners with CIs and listeners with normal hearing listening to vocoded stimuli used native speakers of Mandarin Chinese, which is a tonal language. An advantage of Mandarin is that it contains words that vary in meaning, but differ acoustically only in F0 contour. For example, the syllable /ma/ can have one of four different meanings depending on the contour. Under both CI and vocoder processing, the authors varied the number of channels from 1 to 20, and asked listeners to perform a tone recognition task. In CI patients, they found that performance ranged from near chance (25 percent correct) with only one active electrode to about 70 percent correct with 10 active channels (Wei et al., 2004). Not every CI patient could perform the task, however, as one patient never achieved better than chance performance. In vocoder simulations, tone recognition was about 70 percent correct even with a single channel. These data indicate that it is at least possible for F0 variation to be used as a cue by some CI patients.

However, these data were collected in quiet and it is unclear whether, with current CI and hearing aid technologies, the F0 cue can be as salient in background noise. A considerable amount of attention has been paid over the years to developing ways of providing F0 to listeners who stand to benefit. These efforts are outlined in the next section.

3. Providing F0 as a cue for listeners with impairment

As we have seen, F0 is important for speech understanding, particularly in difficult listening conditions, such as in the presence of a competing background. However, the individuals who perform the poorest in background noise and thus stand to benefit the most from F0 (i.e., listeners with HI and users of CIs) are often the ones who are least able to take advantage of the cue. As a result, over the years many researchers have explored ways of providing F0 information, in addition to other cues, to these listeners.

3.1. Listeners with HI

Various studies have examined the efficacy of delivering F0 to listeners with HI as an aid to lipreading. One possible advantage to providing an F0 cue as opposed to simply amplifying speech is that the narrow-band nature of the F0 cue (which is typically delivered using a pure-tone carrier) makes it possible for higher sound pressure levels to be achieved for a given amount of gain, because all of the amplification can be concentrated in the frequency region in which F0 occurs.

In one study, lipreading performance by profoundly deafened adults was measured either alone, or augmented by acoustically delivered F0, which was extracted using a laryngograph (Rosen et al., 1981). Subjects practiced with another participant over a number of sessions on a connected discourse task. Each partner was in an adjoining sound booth, with a window between so they could see each other. The job of the participant with HI was to read the lips of his or her partner, and repeat what they were saying. The talker could receive auditory feedback from the receiver, but not vice-versa, so that if there was an error in communication, the receiver could indicate this to the talker and they could try again. Performance was measured in the number of words accurately repeated per minute. While only small improvements to lipreading speed were observed over time without the F0 cue, the rate of lipreading increased consistently over time in the presence of F0, and by the end of testing, was shown to be faster than lipreading alone by as much as 2–3 times (Rosen et al., 1981).

Others have examined similar procedures, using F0 extracted from the acoustic signal, rather than with a laryngograph. Although less accurate, this type of extraction is a more realistic implementation, since laryngographs are somewhat infeasible for everyday interactions. For example, when lipreading performance was measured in listeners with normal hearing, the addition of F0 information provided 50 percentage points of enhancement over lipreading alone (Kishon-Rabin et al., 1996). These researchers also examined the efficacy of delivering the F0 cue to individuals with HI using a tactile display (Kishon-Rabin et al., 1996). Users wore the device on the forearm, and it delivered F0 variation via both vibro-tactile (the rate of a pulse train used to vibrate the device was modulated in frequency), as well as spatial-tactile means (the pulse train was delivered via one of 16 solenoids, arranged linearly on the device). In this way, users received both a rate cue, in the form of the modulation rate of the pulse train, and a place cue, conveyed as the place of stimulation on the forearm. Results showed that the device provided an average of 11 percentage points of benefit over lipreading alone. The authors ruled out the accuracy of the extraction method as a limiting factor, and thus concluded that there were limitations inherent in the tactile delivery method that were responsible for the modest enhancements they observed with the device (Kishon-Rabin et al., 1996).

It has also been shown that in addition to F0 variation, voicing and low-frequency amplitude envelope information can provide benefit as cues to lip reading (Grant et al., 1985). Along these lines, another study examined the benefits to lipreading from a wearable aid that presented both F0 and amplitude envelope information acoustically to profoundly deafened individuals (Faulkner et al., 1992). The aid was shown to provide more benefit than simple amplification for 3 of 5 subjects tested in both consonant identification and connected discourse tracking.

3.2. CI patients

Providing pitch information to CI users is a difficult problem. In designing the devices, tradeoffs have had to be made and as we have seen, one tradeoff has been the near elimination of fine-structure cues in favor of envelope cues. As a result, most pitch cues are unusable for most CI patients. There are several reasons for this. First, there is a limit to the number of useful concurrent electrodes. Of course, more electrodes can be added to the electrode array, but they will not necessarily benefit the user. This is because of a phenomenon known as current spread. It is difficult to control the electrical current once it is discharged from the electrode, and a broad region of the basilar membrane is usually stimulated. Thus, if two electrodes are very close in proximity, the user will not be able to distinguish which electrode was stimulated. As a result, there are only a certain number of effective channels that can be used along the length of the basilar membrane. When one considers that the electrodes are fixed in place, it becomes clear that only very gross frequency resolution is possible using standard CI configurations.

Numerous attempts have been made over the years to provide an improved representation of F0 in CIs. Early devices used various techniques to deliver F0, as well as F1 and F2 cues to patients fitted with a CI (Blamey et al., 1984, 1987; Clark et al., 1987; Dowell et al., 1985, 1986; Seligman et al., 1984). One technique utilizes what is known as rate pitch. Typically, the train of electrical pulses delivered to each electrode has a fixed rate. When F0 is represented by rate pitch in a CI, the pulse train of an electrode is modulated in frequency, such that the rate of the electric pulses varies with changes in F0. Another method can be described as place pitch. This method uses a group of contiguous electrodes, and changes in frequency are coded by changes in the electrode being discharged. Early devices demonstrated significant benefits in speech reception by using a combination of these two methods. For example, some early CIs were designed to deliver F0 in the form of rate pitch (Blamey et al., 1984; Clark et al., 1987; Dowell et al., 1985), and the higher ordered formants (F1, F2) in the form of place pitch (Blamey et al., 1987; Clark et al., 1987; Dowell et al., 1986; Seligman et al., 1984). Results showed that speech reception was best when all three formants were represented, rather than either F0 alone, or F0 and F2 (Blamey et al., 1984).

An interesting place-pitch strategy has been employed recently to encode F0 with some success (Geurts and Wouters, 2004). This technique involves delivering current to a pair of adjacent electrodes, and weighting the current to each. If the current to both electrodes is weighted equally, then the pitch percept will be between those obtained when either electrode is stimulated alone (McDermott and McKay, 1994). As the weighting shifts to the more basal electrode, for example, the pitch percept will shift correspondingly higher. This type of stimulation has been called current steering (Koch et al., 2007) or virtual channels (Wilson et al., 2004).

Another place-pitch implementation is known as current focusing, or tripolar stimulation (Miyoshi et al., 1996). This strategy is an attempt to address the problem of current spread. Three adjacent electrodes are used, with the two flanking electrodes carrying the same current load as the center electrode, but reduced in amplitude, and inverted in phase. The idea is that the phase-inverted current from the flanking electrodes will cancel the current from the center electrode as it spreads laterally, thus 'focusing' the current. An examination of cortical activity has confirmed a narrowed current field from tripolar stimulation (Bierer and Middlebrooks, 2002), and better spectral resolution than standard (monopolar) stimulation strategies (Berenstein et al., 2008) has also been reported. However, there has been little, if any, improvement to speech intelligibility as a result of this strategy reported thus far (Mens and Berenstein, 2005). The nature of tripolar stimulation

seems to make it a potentially useful mechanism for delivering a pitch cue to CI patients, since in theory, it would be possible to dynamically adjust the current to the two flanking electrodes to 'steer' the current from the center electrode according to changes in the F0 cue. However, the efficacy of tripolar stimulation in this regard remains to be seen.

Finally, various attempts have been made to improve the F0 representation in CIs in the form of envelope pitch, in which the electric pulses are amplitude-modulated with an extracted F0 (Green et al., 2004, 2005; Milczynski et al., 2009). While improvements have been shown for pitch ranking and melody recognition (Laneau et al., 2004; Milczynski et al., 2009), there has been limited benefit to speech reception reported thus far (Green et al., 2005).

The most common limitation of the wide-spread adoption of F0-based solutions may be the problem of feature extraction (including F0) in noise. Although this is an issue that has received a great deal of attention, and many signal processing refinements have been made over the years to improve the accuracy of feature extraction algorithms, a robust solution has not yet been found.

3.3. EAS patients

We have established that the limited representation of F0 by modern CIs results in difficulty by CI users with speech understanding in the presence of background noise. While much work has been devoted to finding ways to improve F0 representation by the devices, perhaps the best way to deliver F0 to those CI patients who retain some residual hearing is acoustically. For both simulated and real implant processing, the addition of low-frequency acoustic stimulation often enhances speech understanding, particularly when listening to speech in the presence of competing speech (Dorman et al., 2005; Kong et al., 2005; Turner et al., 2004). The benefit of this so-called electric–acoustic stimulation (EAS) occurs even when the acoustic stimulation alone provides little or no intelligibility (i.e., no words correctly identified).

Several recent papers have demonstrated that CI patients may need residual hearing only up to about 125 Hz or so to achieve an EAS benefit (Cullington and Zeng, in press; Zhang et al., in press). In one study for example, EAS patients listened to both sentences in noise and monosyllables in quiet (Zhang et al., in press). The acoustically-delivered stimuli were low-pass filtered at 750, 500, 250, or 125 Hz, or presented broadband and filtered by each listener's audiometric configuration. Results showed that the majority of the EAS benefit was derived from information present below 125 Hz. With monosyllables, for example, scores in the electric-only, EAS-125, and EAS-broadband conditions were 56, 78, and 88 percent correct (Zhang et al., in press). Given that the talker was a male with a mean F0 of 123 Hz, it seems clear that F0 provided the majority of the EAS benefit observed in this study.

Segregation has been suggested as a possible mechanism for the benefits of EAS in noise (Chang et al., 2006; Kong et al., 2005; Qin and Oxenham, 2006). According to this theory, listeners are able to combine the relatively weak pitch information conveyed by the electric stimulation with the stronger pitch cue from the low-frequency acoustic region to segregate target and masker. This is a reasonable hypothesis, because there are results in other contexts that show that F0 aids in the segregation of competing talkers (Assmann, 1999; Assmann and Summerfield, 1990; Bird and Darwin, 1997; Brokx and Nooteboom, 1982; Culling and Darwin, 1993a). Recent reports (Chang et al., 2006; Qin and Oxenham, 2006) have provided evidence that F0 is likely to play an important role independent of any role that the first formant may play. For example, significant benefit was observed when vocoder stimulation was supplemented by 300-Hz low-pass speech (Chang et al., 2006; Qin and Oxenham, 2006). This seems to suggest that F0 is impor-

tant, since it is very unlikely that the first formant is present below 300 Hz (Hillenbrand et al., 1995).

The majority of the work that has been published on this topic points to F0 as a useful cue under EAS conditions. Much of the evidence to support F0 has been rather indirect, however, and there is not complete agreement as to the importance of F0. For example, the first published work that directly examined the contributions of F0 did so under simulated EAS conditions (Kong and Carlyon, 2007), and showed that F0 provided no benefit. These researchers extracted the dynamic changes in F0 from target speech, and used the F0 information to frequency modulate a harmonic complex having a mean F0 equal to the target talker's mean F0. The complex was then combined with a target-plus-background mixture processed through a vocoder which simulated electric stimulation. The rationale was that if an EAS benefit is observed using the modulated complex, it can be concluded that F0 is a useful cue. The authors also examined the independent contributions of the voicing cue (by turning an unmodulated complex on and off with voicing) and the amplitude envelope (by amplitude modulating the complex with the amplitude envelope of the target speech), and found that while both of those cues provided benefit over vocoder-only stimulation, F0 did not. They concluded that because the F0 cue was not shown to be beneficial in their study, segregation was likely not the mechanism for the benefits of EAS. Instead, they suggested glimpsing as an explanation, wherein the low-frequency stimulus (target speech or the amplitude-modulated harmonic complex) provides an indication of when to listen in the vocoder region. Because of comodulation inherent in speech, when the level of the target is high in the low-frequency region, there is a greater likelihood that the signal-to-noise ratio is more favorable in the vocoder region.

A similar experimental design was employed by Brown and Bacon (2009b) using a pure-tone carrier, as opposed to a harmonic complex. These simulation data showed that each of the three cues tested (voicing, amplitude envelope, and F0) contributed a significant amount of benefit under simulated EAS conditions, and that combining all three produced the greatest benefit. The amount of benefit to speech intelligibility was between 24 and 57 percentage points over vocoder-only stimulation, depending on the target and masker materials used. In particular, the tone provided more benefit when the sentence context was high. Fig. 4 shows data collected using a procedure that was identical to that used by Brown and Bacon (2009b), but with high-context sentence materials. Each plot represents percent correct word recognition for CUNY sentences, spoken by a male talker, in either a different male background (squares) or a female background (circles). Processing conditions are depicted along the *x* axis, and include vocoder only (*V*), vocoder plus 500-Hz low-pass target speech (*V*/500), vocoder plus a tone with a frequency equal to the mean F0 of the target talker that was turned on and off with voicing (*V*/T_{voice}), vocoder plus a tone modulated in frequency with the target talker's F0 (*V*/T_{F0}), vocoder plus a tone modulated in amplitude with the amplitude envelope of the low-pass target speech (*V*/T_{env}), and vocoder plus a tone carrying all three cues (*V*/T_{F0-env}). Note that the voicing cue is present in all of the tone conditions. The values to the right of each plot are the amounts of improvement from the frequency- and amplitude-modulated tone in each background over vocoder-only performance. Although the benefit due to F0 and the amplitude envelope is greater here with the high-context sentences than we observed previously with low-context sentences (Brown and Bacon, 2009b), the pattern of results is the same: both F0 and the amplitude envelope contribute nearly equivalent amounts of benefit over vocoder only, and the combination of the two cues provides more benefit than either cue alone.

The effectiveness of the tone has also been confirmed with EAS patients (Brown and Bacon, 2009a), who showed an average benefit
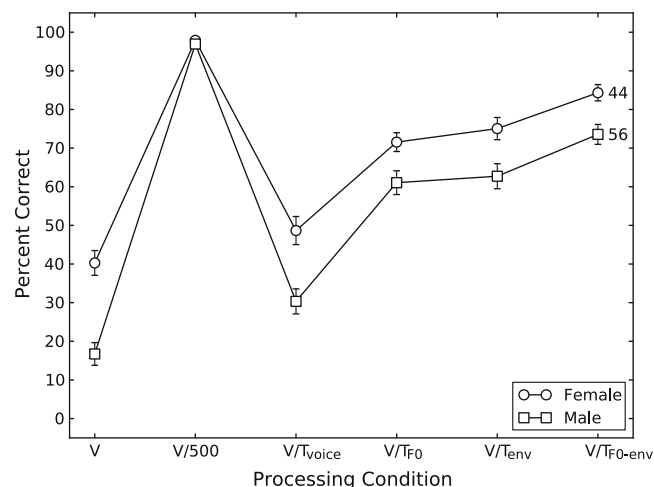
Fig. 4. Mean percent correct intelligibility. Squares represent performance in a male background (different talker from the target), circles represent a female background. Processing conditions are depicted along the *x* axis, and include vocoder only (*V*), vocoder plus 500-Hz low-pass target speech (*V*/500), vocoder plus a tone with a frequency equal to the mean F0 of the target talker that was turned on and off with voicing (*V*/T_{voice}), vocoder plus a tone modulated in frequency with the target talker's F0 (*V*/T_{F0}), vocoder plus a tone modulated in amplitude with the amplitude envelope of the low-pass target speech (*V*/T_{env}), and vocoder plus a tone carrying all three cues (*V*/T_{F0-env}). Note that the voicing cue is present in all of the tone conditions. The values to the right of each plot are the amounts of improvement from the frequency- and amplitude-modulated tone (*V*/T_{F0-env}) in each background over vocoder-only performance (*V*).

over electric-only stimulation of 46 percentage points when the frequency- and amplitude-modulated tone was presented acoustically, as compared to 55 percentage points when the acoustic stimulus was target speech. It has also been reported that the presence of a second tone that carried F0 and amplitude envelope cues from the masker sentence had little to no effect on performance (Brown and Bacon, 2009b), and that the F0 and amplitude envelope cues of the target speech can be carried by a tone much lower in frequency than the mean F0 of the target talker with no loss in the amount of EAS benefit (Brown et al., in preparation). This result in particular is strong evidence against segregation as an explanation for EAS. If listeners are indeed segregating target and masker by combining the strong F0 cue in the acoustic stimulation with the relatively weak pitch cue from the electric stimulation, then altering the pitch cue by shifting it down in frequency should reduce or eliminate the benefit. But these researchers showed that even when the F0 variation from a female target talker whose mean F0 was 213 Hz was applied to a tone with a frequency of 113 Hz, there was no decrease in the benefit due to the tone (Brown et al., in preparation). As a result, the authors have argued against segregation as an explanation for the benefits of EAS, and support the glimpsing account proposed previously (Kong and Carlyon, 2007; Li and Loizou, 2008).

These results seem to indicate that it may be possible to provide F0 (and the amplitude envelope) to EAS patients who have very little residual hearing. For example, if a CI patient has residual hearing up to about 100 Hz, no EAS benefit typically may be observed, even with amplification. One reason for this lack of benefit is that the F0 of most talkers, particularly women and children (Peterson and Barney, 1952), would be too high in frequency to be audible. However, the benefits of EAS may be possible if F0 and amplitude envelope cues are applied to a tone that has been shifted down into a frequency region of audibility (Brown et al., in preparation).

4. Conclusion

Complex waveforms such as speech contain, to a greater or lesser degree, both envelope information and temporal fine-structure

information. In favorable listening situations, providing envelope cues only is often sufficient for good intelligibility. This point is illustrated clearly by the success of the modern CI, which is well-known for discarding most fine-structure information in favor of amplitude envelope information. Users of CIs often perform quite well on speech intelligibility tasks in quiet (Dorman and Loizou, 1997, 1998). On the other hand, speech intelligibility by patients with CIs is typically poor in the presence of background noise (Nelson et al., 2003; Stickney et al., 2004). Indeed, background noise is a significant problem for listeners with HI and CI users, and these difficulties have been linked to reduced fine-structure processing, a common characteristic in both listeners with HI (Hall and Wood, 1984; Lorenzi et al., 2006; Moore et al., 2006), and those fitted with a CI (Fu and Nogaki, 2005; Stickney et al., 2007).

An important consequence of reduced fine-structure processing is an inability to glimpse a target in a fluctuating masker, which leads to poor performance on speech reception tasks in modulated noise (Lorenzi et al., 2006, 2009). Although F0 has been shown to be important for speech understanding in a variety of settings, the cue seems to be particularly useful in difficult listening conditions in which fine-structure cues may be limited, such as those experienced by listeners with HI and users of CIs. Several researchers have cited an inadequate representation of F0 as one of the causes of the reduced speech reception often observed in these impaired populations (Festen and Plomp, 1990; Nelson and Jin, 2004; Qin and Oxenham, 2003, 2005; Stickney et al., 2004, 2007; Summers and Leek, 1998), and although numerous attempts have been made to provide them with the F0 cue, much more work needs to be done.

Acknowledgments

References

Assmann, P.F., 1999. Fundamental frequency and the intelligibility of competing voices. Proc. 14th Int. Cong. Phonetic Sci., 179–182.

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

Bacon, S.P., Opie, J.M., Montoya, D.Y., 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. J. Speech Lang. Hear. Res. 41, 549–563.

Berenstein, C.K., Mens, L.H.M., Mulder, J.J.S., Vanpoucke, F.J., 2008. Current steering and current focusing in cochlear implants: comparison of monopolar, tripolar, and virtual channel electrode configurations. Ear Hear. 29, 250–260.

Bierer, J.A., Middlebrooks, J.C., 2002. Auditory cortical images of cochlear-implant stimuli: dependence on electrode configuration. J. Neurophysiol. 87, 478–492.

Binns, C., Culling, J.F., 2007. The role of fundamental frequency contours in the perception of speech against interfering speech. J. Acoust. Soc. Am. 122, 1765.

Bird, J., Darwin, C.J., 1997. Effects of a difference in fundamental frequency in separating two sentences. Paper for the 11th Int. Conf. on Hear.

Blamey, P.J., Dowell, R.C., Brown, A.M., Clark, G.M., Seligman, P.M., 1987. Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. J. Acoust. Soc. Am. 82, 48–57.

Blamey, P.J., Dowell, R.C., Tong, Y.C., Brown, A.M., Luscombe, S.M., Clark, G.M., 1984. Speech processing studies using an acoustic model of a multiple-channel cochlear implant. J. Acoust. Soc. Am. 76, 104–110.

Brokx, J., Nooteboom, S., 1982. Intonation and the perceptual separation of simultaneous voices. J. Phonetics 10, 23–36.

Brown, C.A., Bacon, S.P., 2009a. Achieving electric–acoustic benefit with a modulated tone. Ear Hear. 30, 489–493.

Brown, C.A., Bacon, S.P., 2009b. Low-frequency speech cues and simulated electric–acoustic hearing. J. Acoust. Soc. Am. 125, 1658–1665.

Brown, C.A., Scherrer, N.M., Bacon, S.P., in preparation. Shifting fundamental frequency in simulated electric–acoustic listening. J. Acoust. Soc. Am.

Carhart, R., Tillman, T.W., 1970. Interaction of competing speech signals with hearing losses. Arch. Otolaryngol. 91, 273–279.

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

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

Chatterjee, M., Peng, S., 2008. Processing F0 with cochlear implants: modulation frequency discrimination and speech intonation recognition. Hear. Res. 235, 143–156.

Clark, G.M., Blamey, P.J., Brown, A.M., Gusby, P.A., Dowell, R.C., Franz, B.K., Pyman, B.C., Shepherd, R.K., Tong, Y.C., Webb, R.L., et al., 1987. The University of Melbourne – nucleus multi-electrode cochlear implant. Adv. Otorhinolaryngol. 38, 1–181. V–IX.

Culling, J.F., Darwin, C.J., 1993a. The role of timbre in the segregation of simultaneous voices with intersecting F0 contours. Percept. Psychophys. 54, 303–309.

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

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

Cullington, H.E., Zeng, F., in press. Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. Ear Hear.

Dorman, M.F., Loizou, P.C., 1997. Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. Am. J. Otol. 18, S113–114.

Dorman, M.F., Loizou, P.C., 1998. The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels. Ear Hear. 19, 162–166.

Dorman, M.F., Loizou, P.C., Rainey, D., 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Am. 102, 2403–2411.

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

Dowell, R.C., Martin, L.F., Clark, G.M., Brown, A.M., 1985. Results of a preliminary clinical trial on a multiple channel cochlear prosthesis. Ann. Otol. Rhinol. Laryngol. 94, 244–250.

Dowell, R.C., Mecklenburg, D.J., Clark, G.M., 1986. Speech recognition for 40 patients receiving multichannel cochlear implants. Arch. Otolaryngol. Head Neck Surg. 112, 1054–1059.

Duquesnoy, A.J., 1983. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. J. Acoust. Soc. Am. 74, 739–743.

Faulkner, A., Rosen, S., 1999. Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio–visual and auditory speech perception. J. Acoust. Soc. Am. 106, 2063–2073.

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

Festen, J.M., Plomp, R., 1983. Relations between auditory functions in impaired hearing. J. Acoust. Soc. Am. 73, 652–662.

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

Fu, Q., Nogaki, G., 2005. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J. Assoc. Res. Otolaryngol. 6, 19–27.

Fu, Q., Chinchilla, S., Nogaki, G., Galvin, J.J.3., 2005. Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. J. Acoust. Soc. Am. 118, 1711–1718.

Fu, Q.J., Shannon, R.V., Wang, X., 1998. Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. J. Acoust. Soc. Am. 104, 3586–3596.

Geurts, L., Wouters, J., 2004. Better place-coding of the fundamental frequency in cochlear implants. J. Acoust. Soc. Am. 115, 844–852.

Grant, K.W., Ardell, L.H., Kuhl, P.K., Sparks, D.W., 1985. The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. J. Acoust. Soc. Am. 77, 671–677.

Green, T., Faulkner, A., Rosen, S., 2004. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J. Acoust. Soc. Am. 116, 2298–2310.

Green, T., Faulkner, A., Rosen, S., Macherey, O., 2005. Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification. J. Acoust. Soc. Am. 118, 375–385.

Hall, J.W., Wood, E.J., 1984. Stimulus duration and frequency discrimination for normal-hearing and hearing-impaired subjects. J. Speech Hear. Res. 27, 252–256.

Healy, E.W., Steinbach, H.M., 2007. The effect of smoothing filter slope and spectral frequency on temporal speech information. J. Acoust. Soc. Am. 121, 1177–1181.

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

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

Hopkins, K., Moore, B.C.J., 2007. Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. J. Acoust. Soc. Am. 122, 1055–1068.

Hopkins, K., Moore, B.C.J., Stone, M.A., 2008. Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. J. Acoust. Soc. Am. 123, 1140–1153.

Houtgast, T., Festen, J.M., 2008. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. Int. J. Audiol. 47, 287–295.

Kishon-Rabin, L., Boothroyd, A., Hanin, L., 1996. Speechreading enhancement: a comparison of spatial-tactile display of voice fundamental frequency (F0) with auditory F0. J. Acoust. Soc. Am. 100, 593–602.

Koch, D.B., Downing, M., Osberger, M.J., Litvak, L., 2007. Using current steering to increase spectral resolution in CII and HiRes 90K users. Ear Hear. 28, 38S–41S.

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

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

Laneau, J., Wouters, J., Moonen, M., 2004. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. J. Acoust. Soc. Am. 116, 3606–3619.

Li, N., Loizou, P.C., 2008. A glimpsing account for the benefit of simulated combined acoustic and electric hearing. J. Acoust. Soc. Am. 123, 2287–2294.

Liu, S., Del Rio, E., Bradlow, A.R., Zeng, F., 2004. Clear speech perception in acoustic and electric hearing. J. Acoust. Soc. Am. 116, 2374–2383.

Loizou, P.C., 1998. Mimicking the human ear. IEEE Sig. Proc. Mag., 101–129.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., Moore, B.C.J., 2009. Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. J. Acoust. Soc. Am. 125, 27–30.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc. Natl. Acad. Sci. USA 103, 18866–18869.

McDermott, H.J., McKay, C.M., 1994. Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. J. Acoust. Soc. Am. 96, 155–162.

Mens, L.H.M., Berenstein, C.K., 2005. Speech perception with mono- and quadrupolar electrode configurations: a crossover study. Otol. Neurotol. 26, 957–964.

Milczynski, M., Wouters, J., van Wieringen, A., 2009. Improved fundamental frequency coding in cochlear implant signal processing. J. Acoust. Soc. Am. 125, 2260–2271.

Miller, G.A., Licklider, J.C.R., 1950. The intelligibility of interrupted speech. J. Acoust. Soc. Am. 22, 167–173.

Miyoshi, S., Iida, Y., Shimizu, S., Matsushima, J., Ifukube, T., 1996. Proposal of a new auditory nerve stimulation method for cochlear prosthesis. Artif. Organs 20, 941–946.

Moore, B.C.J., 2008. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J. Assoc. Res. Otolaryngol. 9, 399–406.

Moore, B.C.J., Glasberg, B.R., Hopkins, K., 2006. Frequency discrimination of complex tones by hearing-impaired subjects: evidence for loss of ability to use temporal fine structure. Hear. Res. 222, 16–27.

Nelson, P.B., Jin, S., 2004. Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 115, 2286–2294.

Nelson, P.B., Jin, S., Carney, A.E., Nelson, D.A., 2003. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 113, 961–968.

Oxenham, A.J., Simonson, A.M., 2009. Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. J. Acoust. Soc. Am. 125, 457–468.

Peters, R.W., Moore, B.C.J., Baer, T., 1998. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J. Acoust. Soc. Am. 103, 577–587.

Peterson, G.E., Barney, H.L., 1952. Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175–184.

Plomp, R., 1978. Auditory handicap of hearing impairment and the limited benefit of hearing aids. J. Acoust. Soc. Am. 63, 533–549.

Plomp, R., 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. J. Speech Hear. Res. 29, 146–154.

Plomp, R., 1994. Noise, amplification, and compression: considerations of three main issues in hearing aid design. Ear Hear. 15, 2–12.

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

Qin, M.K., Oxenham, A.J., 2005. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. Ear Hear. 26, 451–460.

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

Rosen, S.M., Fourcin, A.J., Moore, B.C.J., 1981. Voice pitch as an aid to lipreading. Nature 291, 150–152.

Seligman, P.M., Patrick, J.F., Tong, Y.C., Clark, G.M., Dowell, R.C., Crosby, P.A., 1984. A signal processor for a multiple-electrode hearing prosthesis. Acta Otolaryngol. Suppl. 411, 135–139.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Spitzer, S.M., Liss, J.M., Mattys, S.L., 2007. Acoustic cues to lexical segmentation: a study of resynthesized speech. J. Acoust. Soc. Am. 122, 3678–3687.

Stickney, G.S., Assmann, P.F., Chang, J., Zeng, F., 2007. Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. J. Acoust. Soc. Am. 122, 1069–1078.

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

Stone, M.A., Füllgrabe, C., Moore, B.C.J., 2008. Benefit of high-rate envelope cues in vocoder processing: effect of number of channels and spectral region. J. Acoust. Soc. Am. 124, 2272–2282.

Summers, V., Leek, M.R., 1998. F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. J. Speech Lang. Hear. Res. 41, 1294–1306.

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

Wei, C., Cao, K., Zeng, F., 2004. Mandarin tone recognition in cochlear-implant subjects. Hear. Res. 197, 87–95.

Whalen, D.H., Abramson, A.S., Lisker, L., Mody, M., 1993. F0 gives voicing information even with unambiguous voice onset times. J. Acoust. Soc. Am. 93, 2152–2159.

Whitmal, N.A., Poissant, S.F., Freyman, R.L., Helfer, K.S., 2007. Speech intelligibility in cochlear implant simulations: effects of carrier type, interfering noise, and subject experience. J. Acoust. Soc. Am. 122, 2376–2388.

Wilson, B., Sun, X., Schatzer, R., Wolford, R., 2004. Representation of fine structure or fine frequency information with cochlear implants. Int. Cong. Ser. 1273, 3–6.

Zhang, T., Dorman, M.F., Spahr, A.J., in press. 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.