Dual-carrier processing to convey temporal fine structure cues: Implications for cochlear implants

Frédéric Apoux,^{a)} Carla L. Youngdahl, Sarah E. Yoho, and Eric W. Healy

Speech Psychoacoustics Laboratory, Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210, USA

(Received 14 October 2014; revised 22 July 2015; accepted 23 July 2015; published online 14 September 2015)

Speech intelligibility in noise can be degraded by using vocoder processing to alter the temporal fine structure (TFS). Here it is argued that this degradation is not attributable to the loss of speech information potentially present in the TFS. Instead it is proposed that the degradation results from the loss of sound-source segregation information when two or more carriers (i.e., TFS) are substituted with only one as a consequence of vocoder processing. To demonstrate this segregation role, vocoder processing involving two carriers, one for the target and one for the background, was implemented. Because this approach does not preserve the speech TFS, it may be assumed that any improvement in intelligibility can only be a consequence of the preserved carrier duality and associated segregation cues. Three experiments were conducted using this “dual-carrier” approach. All experiments showed substantial sentence intelligibility in noise improvements compared to traditional single-carrier conditions. In several conditions, the improvement was so substantial that intelligibility approximated that for unprocessed speech in noise. A foreseeable and potentially promising implication for the dual-carrier approach involves implementation into cochlear implant speech processors, where it may provide the TFS cues necessary to segregate speech from noise.

© 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4928136>]

[EB]

Pages: 1469–1480

I. INTRODUCTION

In recent years, evidence has accumulated that some of the mechanisms used by the normal auditory system to extract the target speech signal from a sound mixture might rely on information present in the temporal fine structure (TFS). Because it is difficult, if not impossible, to determine the “TFS” of complex signals such as speech, an operational definition of TFS will be used here, corresponding to the cosine of the Hilbert instantaneous phase (see Apoux *et al.*, 2011). Most of the evidence supporting the use of TFS comes from studies showing better intelligibility of natural (i.e., unprocessed) speech relative to speech processed to replace the original TFS with tones or noise bands. This so-called vocoder approach (Shannon *et al.*, 1995) will be referred to as single-carrier processing because it involves only one carrier or TFS in each spectral band. The drop in intelligibility associated with single-carrier processing is generally larger when the background fluctuates over time, suggesting a greater role of TFS in temporally fluctuating backgrounds (e.g., Nelson *et al.*, 2003; Qin and Oxenham, 2003, 2006; Stickney *et al.*, 2005; Füllgrabe *et al.*, 2006; Gnansia *et al.*, 2009). Accordingly, it is believed that TFS cues play a critical role when listening to speech in noise, especially if the noise is fluctuating.

While the detrimental effect of single-carrier vocoder processing on speech recognition in noise is not debatable, there is some controversy regarding how the normal auditory system uses TFS cues to better process speech in noise. Thus

far, most studies on this topic have suggested that TFS provides acoustic speech information. This “speech-information hypothesis” is supported by one finding: normal-hearing (NH) listeners can understand speech when presented with the isolated TFS (Lorenzi *et al.*, 2006).¹ Although widely cited, this hypothesis is somewhat controversial. Indeed, several studies have demonstrated that speech envelope information remains in the isolated TFS and that it is this envelope information that can be recovered and used to understand speech (Ghitza, 2001; Zeng *et al.*, 2004; Gilbert and Lorenzi, 2006; Apoux *et al.*, 2011; Swaminathan *et al.*, 2014).

Adding to this controversy, several studies have demonstrated that NH listeners do not extract much speech information from the TFS when both envelope and TFS are present in the stimulus. In particular, we have conducted a series of experiments in which the signal-to-noise ratio (SNR) of the envelope and that of the TFS were manipulated independently (Apoux and Healy, 2013; Apoux *et al.*, 2013). We showed that varying the SNR of the TFS has little impact on speech intelligibility, suggesting that NH listeners do not rely on information in the TFS to understand speech in “real-world” situations (i.e., when presented with both the envelope and TFS at SNRs in the range from –18 to 12 dB). In contrast, performance was highly correlated with the SNR of the envelope, suggesting that speech intelligibility is primarily supported by information provided by the temporal envelope. Accordingly, we have suggested that the speech-TFS may not contain and/or provide speech information; this is in stark contradiction with the speech-information hypothesis.

^{a)}Electronic mail: fred.apoux@gmail.com

A question that arises at this point is why intelligibility is affected by single-carrier vocoder processing if the TFS does not provide speech information. Recently it has been suggested that TFS cues may be involved in streaming (Qin and Oxenham, 2003; Hopkins and Moore, 2010; Apoux and Healy, 2010, 2011, 2013). In this view, the TFS provides cues that may be used to segregate the target speech from the background. Although this “streaming hypothesis” is more consistent with the larger importance of TFS in fluctuating than in steady backgrounds, it has received limited direct attention until now. As a consequence, there is little direct evidence of a role of TFS cues in streaming. Even the TFS studies suggesting that the speech-TFS does not provide speech information have not provided such direct evidence. The primary goal of the present study was therefore to provide direct empirical evidence that TFS is involved in sound source segregation.

To achieve this goal, we implemented a technique directly inspired by the results of a recent study by Apoux and Healy (2013). In this study, we suggested that the use of a “single TFS” as carrier may be one of the primary factors underlying the detrimental effect of traditional vocoder processing in noise. Indeed one should not be surprised that it is especially challenging to process a stimulus consisting of multiple envelopes from independent sources, all imposed on a single carrier. In an attempt to clarify the consequences of having independent envelopes (i.e., from independent sources) imposed on a single carrier, Apoux and Healy (2013) manipulated the number and nature of carriers used to convey the envelopes of two independent sound sources. These envelopes were mixed and imposed (1) on the isolated target TFS, (2) on the isolated background TFS, or (3) on TFS extracted from the mixed target+background. The results of this study revealed that preserving only the target TFS is equivalent to preserving only the background TFS, suggesting that the nature of the TFS has little influence on streaming (and intelligibility). More importantly, it revealed that preserving the mixed TFSs is by far the most advantageous condition, suggesting that the critical factor for effective speech recognition in noise is the duality of carriers.

Previous work employing multiple carriers has led to limited, if any, improvements in speech recognition. Gnansia *et al.* (2010) examined temporally interleaved vocoded voices. However, syllable recognition was not improved when the carriers for the two voices were slightly shifted in frequency, relative to an unshifted condition in which the two voices were vocoded using a single carrier. Deeks and Carlyon (2004) examined more directly the possibility to increase streaming and speech recognition in noise using multiple carriers. The results, however, were not compelling and led the authors to conclude that “differences in pulse rate are unlikely to prove useful for concurrent sound segregation.” Here we reexamine this issue and argue that the apparent contradiction between this conclusion and the hypotheses tested in the present study may have resulted from the particular approach and the limited number of conditions employed in these previous studies.

Consistent with the finding that speech-TFS likely does not provide speech information, we first hypothesized that it

may not be necessary to provide the original speech fine structure. In other words, introducing synthetic but perceptually relevant cues in the TFS may be sufficient to support effective segregation and therefore substantially improve speech recognition in noise. Consistent with the importance of having multiple carriers, we further postulated that such synthetic cues could result from employing two independent carriers, one conveying the target and one conveying the background envelope. Here these carriers were two trains of pulses that differed only in their nominal rate. Based on previous work suggesting a role of TFS in pitch perception (e.g., Smith *et al.*, 2002), it was finally hypothesized that the rate difference between the carriers used for the target and the background would associate a unique “pitch” percept to each signal; this in turn should result in improved segregation. In contrast to current approaches in which the carriers are manipulated to somehow reflect the TFS of the incoming sounds (e.g., Mitterbacher *et al.*, 2005a; Mitterbacher *et al.*, 2005b), our so-called “dual-carrier” strategy employs arbitrary-rate pulse trains that are not derived from the TFS of the encoded sounds. The use of arbitrary carriers is consistent with our view that the original fine structure is not necessary because the TFS does not provide speech information *per se*. More importantly, we demonstrate that divergent approaches, such as the dual-carrier approach proposed here, can lead to significant improvements in sound-source segregation.

A foreseeable and potentially promising implication for the dual-carrier approach involves implementation into CI speech processors where it may provide the TFS cues necessary for the user to extract speech from noise and improve speech-recognition performance. Such implication is consistent with the suggestion that “improving the ability to use TFS should be a goal for designers of [...] cochlear implants (CIs)” (Lorenzi *et al.*, 2006).

II. EXPERIMENT 1: BENEFIT OF DUAL-CARRIER PROCESSING ON SPEECH INTELLIGIBILITY IN NOISE

The goal of this first experiment was to investigate the role of TFS in streaming by evaluating the benefit of dual-carrier processing. The carrier rates were selected to have no harmonic relationship. We reasoned that limiting harmonic relationships between the target and background carriers would result in greater discriminability and therefore better segregation.² Harmonic relationship was avoided simply by using prime numbers for all rates. Because the effect of TFS has been shown to be larger in modulated backgrounds, sentences were used as the background in this experiment. However, all background sentences were played backward to eliminate linguistic content to some extent and to limit confusions with the target while preserving their speech-like acoustic characteristics (time-reversed speech; TRS).

A. Methods

1. Subjects

Twenty-four NH listeners participated in this experiment (22 females). Their ages ranged from 20 to 28 yr

(average = 21.6 yr). All participants had pure-tone air-conduction thresholds of 20 dB hearing level (HL) or better at octave frequencies from 250 to 8000 Hz (ANSI, 2004, 2010). They were paid an hourly wage or received course credit for their participation. This study was approved by The Ohio State University Institutional Review Board.

2. Speech material and processing

The target stimuli consisted of sentences from the IEEE corpus (IEEE, 1969) produced by a single male talker. Background sentences produced by one male and one female talker were randomly selected from the AzBio set (Spahr et al., 2012). The recordings were digitized with 16-bit precision at a sampling rate of 22 050 Hz. Three conditions were compared. In a first condition, target and background were summed and presented to subjects without further processing: unprocessed condition (UNP). This UNP condition represents the “gold-standard” performance of NH listeners. In the other two conditions, target and background were vocoded independently prior to summation using two independent single-carrier vocoders. Each single-carrier vocoder was implemented as follows: stimuli were filtered into ten contiguous frequency bands ranging from 80 to 7563 Hz using two cascaded 12th-order digital Butterworth filters so that the filtering roughly simulated the number of independent channels typically available in CI users (e.g., Friesen et al., 2001). Stimuli were filtered in both the forward and reverse directions (i.e., zero-phase digital filtering) so that the filtering process produced zero phase distortion (for more details, see Apoux and Healy, 2009). Each band was three ERB_N wide (normal equivalent rectangular bandwidth; Glasberg and Moore, 1990). The envelope was extracted from each band by half-wave rectification and low-pass filtering at cf_m (eighth-order Butterworth, 48 dB/octave roll-off, see Healy and Steinbach, 2007). The value for cf_m was independently computed for each band to equal half the bandwidth in hertz of the ERB_N at the center of the band (Apoux and Bacon, 2008). The filtered envelopes were then used to modulate pulse train carriers having rates ranging from 89 to 367 pulses per second (pps). These rates were chosen to cover (and exceed) the average voice fundamental frequency range for men, women, and children. The individual pulse duration was ten samples (approximately 0.45 ms). The modulated pulse trains were band-pass filtered to restrict their frequency range to the bandwidth of the corresponding channel and summed over all channels to produce the single-carrier stimulus. The output of the two single-carrier vocoders, one for the target and one for the background, were finally summed. The resulting stimuli were a sound mixture that was made up of two modulated pulse trains the rates of which differed by as few as 0 and as much as 238 pps.³

When the difference between the two carrier rates was 0, the processing resulted in the target and background envelopes sharing a single carrier as the carriers were in phase. Accordingly, these conditions will be referred to as single-carrier conditions (SC), and they were used as reference. It should be noted that the current implementation may differ slightly from a traditional SC vocoder in which the complex

envelope is used to modulate a single pulse train. We chose to use two parallel SC vocoders because this implementation is more comparable to dual-carrier processing. When the difference between the two carrier rates was not zero, the processing resulted in the target and background envelopes being conveyed by two independent carriers. Accordingly, these conditions will be referred to as dual-carrier conditions (DC). Six carrier rates were used (109, 157, 197, 241, 283, and 347 pps). All possible combinations of target and background carrier rates were tested, resulting in 36 conditions. Six of these were the SC conditions, one at each rate. The remaining 30 combinations involving different rates for the target and background carriers were DC conditions.

In DC conditions, the difference between target and background carrier rates ranged from 40 to 238 pps. Whereas streaming may not be expected to improve substantially beyond a 238 pps difference, NH subjects may still be able to segregate the target from the background when the difference between the two carrier rates is below 40 pps (e.g., Gaudrain et al., 2007, 2008). To provide insight into the smallest possible rate difference for target/background segregation, four additional background carrier rates were tested in combination with each target carrier rate. These additional background carrier rates were also prime numbers roughly 10 and 20 pps below and above each target carrier rate.⁴ The two additional background carrier rates below and two above each target carrier rate resulted in 24 additional conditions for a total of 60 conditions (see Table I and Fig. 2).

3. Procedure

The 24 subjects were randomly and evenly divided into three groups. Each group completed 21 conditions corresponding to the combination of two target carrier rates and 10 background carrier rates plus UNP. Order of conditions was randomized but blocked by target carrier rate. This was done to provide the necessary familiarity with the target carrier rate so that effects could be observed that are more representative of the potential influence of DC processing. However, conditions were presented in random order within each target carrier rate set. The target carrier rate that was first completed was balanced within each group.

Subjects were tested individually in a double-walled, sound-attenuated booth. Stimuli were presented diotically through Sennheiser HD 280 Pro circumaural headphones. The experiments were controlled using custom MATLAB routines running on personal computers equipped with

TABLE I. The four additional background carrier rates (pps) tested with each target carrier rate and roughly corresponding to −20, −10, +10, and +20 pps.

−20	−10	Target	+10	+20
89	101	109	113	127
139	149	157	167	179
179	191	197	199	211
223	233	241	251	263
263	271	283	293	307
321	337	347	359	367

FIG. 1. Benefit in percentage points of dual-carrier processing (DC) relative to single-carrier processing (SC) as a function of the target carrier rate. The parameter is the background carrier rate. The dashed line indicates the unprocessed condition (UNP) score in percentage points relative to SC.

high-quality digital-to-analog converters (Echo Gina3G). Subjects were instructed to repeat as many words as possible. A short practice was provided prior to data collection consisting of three blocks with each block corresponding to recognition of eight sentences not used during formal testing. The first block was in the UNP condition and the other two blocks were in the SC condition, one at each target rate assigned to the subject. Feedback was not provided in any session. After practice, each subject completed the 21 experimental blocks. Each experimental block corresponded to recognition of 14 sentences randomly selected from the IEEE corpus. There was no replacement so that each sentence was heard only once, and no subject had prior exposure to the sentences. The total duration of testing, including practice, was approximately 90 min. Presentation level was

set to 70 dBA. The SNR was set to +3 dB. This SNR was chosen so that performance in the SC conditions was approximately 15%, allowing for a large UNP-SC difference. It also represents an SNR value that may be commonly encountered in noisy everyday environments.

B. Results and discussion

The primary goal of the present experiment was to provide direct evidence of the contribution of TFS cues to target/background segregation by showing an improvement in speech recognition in noise under DC conditions when compared to SC conditions. Therefore Fig. 1 shows the data in terms of the benefit (in percentage points) following DC processing relative to the performance in SC (group mean percent correct intelligibilities for these conditions are available in Fig. 2). In this figure, a subset of data is shown corresponding to benefit observed with the six target carrier rates when combined with the six background carrier rates that were used in all the target conditions. For reference, a dashed line indicates the score obtained in the UNP condition, also relative to SC. This “benefit” was roughly 70% points on average. It is apparent from Fig. 1 that the benefit achieved with DC processing did not reach the performance observed in the reference UNP condition. However, the increase in intelligibility in noise resulting from synthetic TFS cues was still substantial, averaging 36% points with the largest improvement approaching 50% points. The target carrier rates tested in the present study resulted in a minimum improvement of 24% points with the two most extreme rates, 109 and 347 pps, providing the smallest average benefit. Interestingly, these two values were also the rates associated with the largest average benefit when used as the background carrier rate.

FIG. 2. Mean sentence recognition scores as a function of the background carrier rate. Each panel corresponds to one target carrier rate. The symbols all represent dual-carrier conditions except the single point in each panel where target and background rates coincide. This corresponds to the SC condition, which is generally the lowest point in each curve. Error bars indicate 1 standard error. The group mean score on unprocessed speech in noise (UNP) was 83%.

The present findings are in stark contradiction with the conclusions of Deeks and Carlyon (2004) that rate differences are unlikely to provide consistent cues for segregation. Indeed the present data clearly demonstrate that rate differences can substantially improve speech intelligibility in noise, presumably by providing cues used to segregate the two signals. This contradiction most likely resulted from differences between the approach used here and that employed by Deeks and Carlyon (2004). There were differences between studies with regard to how the pulsatile carriers were prepared. But more importantly, the two studies differed in the conditions tested. Deeks and Carlyon only tested two “rates”: 80 and 140 Hz.⁵ In the present study, a larger number of rates was employed, providing increased opportunity to observe a benefit and an alternate interpretation. While Deeks and Carlyon concluded an absence of effect, they still observed a benefit at the 140 Hz target rate. The absence of effect was only true for the 80 Hz target rate. As mentioned previously, the average benefit in the present experiment 1 was generally lower at the lowest target carrier rate (109 pps), and it is possible that this benefit would further decrease at lower rates. This interpretation is consistent with the results of a preliminary experiment showing no benefit with a target carrier rate of 79 pps, irrespective of the background carrier rate (67 or 167 pps).⁶ Therefore the limited benefit in Deeks and Carlyon may simply reflect a lack of effect for low target rates. Although it is unclear why very low target carrier rates do not provide improvement, the current data do make clear that the lack of benefit is not uniquely attributable to the relative rates of the target and background carriers as suggested by Deeks and Carlyon. Perhaps more importantly, it is not attributable to the DC approach *per se*.

A secondary goal of the present experiment was to investigate the smallest rate difference necessary for target/background segregation. Figure 2 shows sentence recognition as a function of the background carrier rate. Each panel corresponds to one target carrier rate. As expected, intelligibility was lowest (with one exception) when the target and background carrier rate was the same (i.e., SC). It generally increased as the difference between the target and background carrier rates increased. This increase seemed relatively symmetrical with intelligibility gradually improving up to a roughly 50 pps difference with an apparent asymptote by this 50 pps difference in most conditions. In other words, the benefit from DC processing is maximal when the target and background carrier rates are separated by at least 50 pps. This pattern also confirms what was evident in Fig. 1: there was no apparent interaction between the target and the background carrier rates as the functions are reasonably flat outside the dip corresponding to SC. Overall it may be concluded from these data that sentence recognition is mainly influenced by the rate of the target carrier and that the rate of the background carrier has only a very limited effect on intelligibility provided that the two carrier rates differ by at least 50 pps.

A one-way analysis of variance with factor background carrier rate was performed for each target carrier rate. Prior to these analyses, percentage correct scores were subjected to

arcsine transform (Studebaker, 1985). All six analyses yielded a significant effect of background carrier rate [$F(9,70) \geq 4.83$, $p \leq 0.001$]. *Post hoc* comparisons (Bonferroni-corrected *t*-tests) were performed to determine which DC conditions significantly differed from the SC condition. For two functions in Fig. 2 (157 and 241 pps), all DC scores were significantly different from those in SC. For two other functions (109 and 283 pps), all DC scores were significantly different from SC except for the background carrier rate just above SC (i.e., 113 for 109 and 293 for 283). For one function (197 pps), all DC scores were significantly different from SC except for the background carrier rate just below SC (191 pps) and the two just above (199 and 211 pps). Finally, only four DC scores were significantly different from SC in the 347-pps condition (109, 157, 197, and 241 pps).

III. EXPERIMENT 2: BENEFIT OF DC PROCESSING USING HARMONICALLY RELATED CARRIERS

It is well established that the auditory system has a tendency to perceptually fuse spectral components the frequencies of which are multiple integers of the same fundamental frequency (F_0 ; Moore *et al.*, 1986). This effect is partly responsible for the grouping of individual spectral components into an auditory object. In experiment 1, this so-called harmonicity likely played a role in grouping together the spectral components comprising a carrier. In contrast, the harmonic relationships between the different target and background carriers was limited by using prime numbers for the rates, therefore limiting potential interference between target and background.

In experiment 2, rates were chosen to have various degrees of “harmonicity” across carriers (or carriers’ components). Our interest in testing rates that may potentially interfere with each other lay in the fact that harmonic relationship may greatly facilitate the implementation in CIs. A potential pitfall, however, is that at least some spectral components of each carrier are more likely to overlap in the spectral domain, and therefore these components could be attributed to either carrier. From a streaming perspective, this partial overlap may introduce some ambiguity regarding the source of certain components. Accordingly, one may reasonably expect that a partial harmonic relationship between carriers will produce a decrease in intelligibility. The extent of the potential decrease, however, remains unknown, and it may be small enough to still consider an alternative carrier implementation. The potential decrease may also be partially compensated for by the use of a time-based segregation mechanism, as suggested by the glimpsing model (Cooke, 2006; Apoux and Healy, 2009). The primary goal of experiment 2 was therefore to assess the effect of introducing partial or complete harmonic relationships between the target and background carriers.

A. Methods

Data were collected from 24 new NH listeners (22 females). Their ages ranged from 19 to 24 yr (average = 20.6 yr). All had pure-tone air-conduction thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz.

FIG. 3. Same as Fig. 1 but for harmonically related rates.

Experiment 2 differed from experiment 1 only in the carrier rates tested. All other methodological and procedural details were identical. The “primary” rates used in experiment 2 were all multiple integers of 50 pps. As in experiment 1, six primary rates were used (100, 150, 200, 250, 300, and 350 pps), and all combinations of these six values were tested, resulting in 6 SC conditions and 30 DC conditions. Further, four additional background carrier rates were tested in combination with each target carrier rate. These additional background carrier rates were 10 and 20 pps below and above each target carrier rate (80, 90, 110, 120, 130, 140, 160, 170, 180, 190, 210, 220, 230, 240, 260, 270, 280, 290, 310, 320, 330, 340, 360, and 370 pps). These additional background carrier rates resulted in 24 additional conditions for a total of 60 conditions (these conditions are displayed in Fig. 4).

B. Results and discussion

Figure 3 shows the data in terms of the benefit (in percentage points) from DC processing relative to performance

in SC. As in Fig. 1, a subset of data is shown, corresponding to the benefit observed with the six rates common to all target and background carrier conditions. A dashed line indicates the “benefit” obtained in the UNP condition. This benefit was approximately 73% points on average and, consistent with experiment 1, UNP was the most advantageous condition. Not surprisingly, using harmonically related rates in the DC conditions did not help scores match this reference condition. The increase in intelligibility resulting from providing synthetic TFS cues, however, remained quite substantial. It averaged 35% points with the largest improvement approaching 55% points. As in experiment 1, subjects demonstrated increased speech recognition in noise in all DC conditions with the smallest benefit at 19% points. Again the two most extreme target rates provided the smallest benefit on average (26% points). Overall the generally similar benefit observed in experiments 1 and 2 indicates that the harmonic relationship between target and background carrier rates has little influence on the segregation process involved here.

Experiment 1 showed that no further improvement was observed beyond a 50 pps difference between carrier rates. The effect of rate separation was also assessed in experiment 2, and the results of this assessment are shown in Fig. 4. The lowest data point always corresponds to the SC condition. Asymptotic performance was again generally achieved when the difference between the two carrier rates was 50 pps or greater. More interestingly, the six patterns of data were very similar to those observed in experiment 1. This remarkable similarity further supports the idea that the relationship between the target and the background carrier rates has a limited influence on sentence recognition provided that the difference between the two rates is at least 50 pps.

Figure 4 further illustrates the limited influence of the harmonic relationship between the target and background

FIG. 4. Same as Fig. 2 but for harmonically related rates. The group mean score on unprocessed speech in noise (UNP) was 83% as it was in experiment 1.

carrier rates. In the left upper panel, for instance, the function for the 100-pps target rate is fairly linear above 110 pps with no dip (or peak) at 200 and 300 pps. This was also true for target carrier rates of 200 and 300 pps for which no significant dip was observed at 100 pps. This absence of interaction between harmonically related carrier rates is not consistent with a role of harmonicity. One may argue that this is so because the rates tested in experiment 2 already shared a common $F0$ (50 pps) but the similarity between the present data and those in experiment 1, in which all the rates were prime numbers, makes such an interpretation unlikely.

A one-way analysis of variance with factor background carrier rate was again performed for each target carrier rate using arcsine-transformed scores. Five of the six analyses showed a significant effect of background carrier rate [$F(9,70) \geq 5.29, p \leq 0.001$]. The non-significant analysis was obtained for the 350-pps condition. *Post hoc* comparisons (Bonferroni-corrected *t*-tests) were also performed (1) to determine which DC conditions were significantly different from the SC condition and (2) to evaluate the effect of harmonicity. For this latter analysis, performance for each common multiple or divisor of the target carrier rate was compared to that for the background carrier rate directly preceding and following. A dip in performance was defined as a significantly different score from that observed in both adjacent conditions. For three functions in Fig. 4 (100, 150, and 250 pps), all DC scores were significantly different from SC. For one function (200 pps), only one DC score was not significantly different from SC (210 pps). For the last function (300 pps), two DC scores were not significantly different from SC (290 and 310 pps). These analyses did not reveal an effect of harmonicity as not one of the multiples or divisors of the target carrier rate was found to be significantly different from both of its adjacent conditions.

As mentioned previously, DC processing may have direct implications for CI speech coding strategies. Thus far, DC stimuli were created by adding the outputs of two SC vocoders. Such implementation may not be ideal for CIs because it may not be practical to generate two simultaneous pulse trains with completely unrelated rate values. The results of experiment 2 are noteworthy in that they demonstrate how carriers that share a large number of harmonics do not interfere with each other. Consequently, they also suggest the possibility to create DC stimuli using a single unique pulse train as a single unique rate can encode two harmonically related pulse trains. Because the specifics of an implementation in CIs are beyond the scope of the present study, we will only describe briefly how DC stimuli can be created using a unique pulse train. Consider for instance the two carriers with rates of 100 and 200 pps. A common multiple is 200 pps. Therefore one could generate a unique pulse train at 200 pps on each electrode instead of the two required in the implementation of DC processing employed in experiment 1. In this case, each pulse would (1) carry only the target, (2) carry only the background, (3) carry both, or (4) be zeroed. Because the time interval between the pulses carrying the target would be different from that between the pulses carrying the background, a common pulse train implementation would provide the same percept as does two

separate carriers. Further, the two implementations differ only in appearance as the resulting stimuli are strictly identical. While this alternate implementation of DC processing would also be possible with the rates used in experiment 1, it would require the rate of the unique pulse train to be impractically high because of the use of prime numbers.

IV. EXPERIMENT 3: BENEFIT OF DC PROCESSING IN SPEECH-SHAPED NOISE

As mentioned previously, the drop in intelligibility associated with SC vocoder processing is generally larger when the background fluctuates over time, suggesting a greater role of TFS in temporally fluctuating backgrounds. It is possible then that the benefit from DC processing might be limited in steady backgrounds such as speech-shaped noise (SSN). The purpose of experiment 3 was to assess the benefit from DC processing in a steady background. Two SNR values were used in the present experiment. One was chosen to equate the SNR value to that of the TRS background in experiment 2, and the other was chosen to approximately equate SC performance to that observed in the TRS background.

A. Methods

Data were collected from 32 NH listeners (29 females). Their ages ranged from 19 to 25 yr (average = 21.1 yr). Two subjects had previously participated in experiment 2 (novel sentences were used). All had pure-tone air-conduction thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz with the exception of one subject who had thresholds of 25 dB at 250 and 500 Hz in the right ear. All methodological and procedural details in experiment 3 were identical to those in the previous experiments. The same six rates used in experiment 2 were used here (100, 150, 200, 250, 300, and 350 pps), and all combinations of these six values were tested, resulting in 6 SC conditions and 30 DC conditions. No additional conditions were tested in experiment 3. The SNRs were -3 and $+3$ dB. The subjects were randomly and evenly divided into four groups with each group completing 19 conditions (3 target rates $\times$ 6 background rates plus UNP) at one SNR.

B. Results and discussion

Figure 5 shows the benefit (in percentage points) from DC processing relative to the performance in SC. The top and bottom panels show the -3 and $+3$ dB SNR data, respectively. The average performance in SC was 16% and 42% at -3 and $+3$ dB SNR, respectively. In each panel, a dashed line indicates the score obtained in the UNP condition relative to SC. This “benefit” was approximately 40% points at -3 dB and 48% points at $+3$ dB. Thus the difference between UNP and SC was smaller than in experiments 1 and 2. This is not surprising considering that the reduction in performance resulting from SC vocoder processing is generally smaller in steady backgrounds. One expected consequence is that the benefit from DC processing was inevitably limited, averaging approximately 22% points at -3 dB and 20% points at $+3$ dB. While the raw benefit was comparable across SNRs,

FIG. 5. Each panel shows the benefit in percentage points of DC condition relative to SC condition as a function of the target carrier rate. The parameter is the background carrier rate. The dashed line indicates the “benefit” in UNP. The upper and lower panels reflect scores in speech-shaped noise at -3 and $+3$ dB SNR, respectively.

the largest benefit was observed at -3 dB. At this lowest SNR, DC processing allowed subjects to come within 10% points of unprocessed speech in noise (i.e., UNP, “gold standard”) in 6 of the 25 DC conditions. In other words, DC processing allowed subjects to close 3/4 of the VOC-to-UNP gap. However, a perhaps more remarkable result is that DC processing allowed subjects to match or even exceed their performance in UNP in two other conditions.

A one-way analysis of variance with factor background carrier rate was performed for each target carrier rate in both SNR conditions using arcsine transformed scores. In the -3 -dB SNR condition, all six analyses showed a significant effect of background carrier rate [$F(5,42) \geq 3.22$, $p \leq 0.015$]. In the $+3$ -dB-SNR condition, five of the six analyses showed a significant effect of background carrier rate [$F(5,42) \geq 3.31$, $p \leq 0.013$]. The non-significant analysis was obtained for the 200-pps condition. *Post hoc* comparisons (Bonferroni-corrected *t*-tests) were also performed to determine which DC conditions were significantly different from the SC condition. In the -3 -dB SNR condition, the benefit of DC processing was found to be significant for all conditions at the four lowest target rates (100–250 pps). For the two highest target rates (300 and 350 pps), a significant effect was found for two and three background rates, respectively. In the $+3$ -dB SNR condition, the results were slightly more complex. For the 100 pps target rate, all comparisons

were significant. For the 250 and 350 pps target rates, all comparisons were significant except that against the 150 pps background rate. For the 150 pps target rate, all comparisons were significant except that against the 250 pps background rate. For the 300 pps target rate, a significant effect was found for three background rates. Overall these statistical analyses suggest that DC processing can provide significant intelligibility improvement in steady backgrounds especially at low SNRs.

V. GENERAL DISCUSSION

A. Role of TFS in streaming

The primary goal of the present study was to provide direct evidence of the role of TFS in sound-source segregation. This goal was achieved by implementing a technique in which the TFS potentially provides synthetic cues for streaming while conveying no speech information. Speech intelligibility in noise improved substantially when these synthetic cues were introduced to the sound mixture. The improvement was numerically, but not necessarily relatively, larger in fluctuating than in steady backgrounds. This is because the effect of removing TFS cues from the sound mixture (the UNP-SC difference) is larger in fluctuating than in steady backgrounds, hence limiting the potential improvement in steady backgrounds. Overall these findings are consistent with a primary role of TFS in sound-source segregation. First, they demonstrate that the drop in intelligibility associated with SC vocoder processing can be largely offset without reintroducing any speech information in the TFS. Second, the present findings show that the drop in intelligibility associated with SC vocoder processing can be largely offset by reintroducing in the TFS some of the cues associated with streaming. Taken together, the present results and previous work showing that TFS cues provide limited, if any, speech information (Apoux and Healy, 2013; Apoux *et al.*, 2013) strongly suggest that the primary role of TFS is to support sound-source segregation.

One of the mechanisms by which TFS can support sound-source segregation is by allowing, at each moment in time, the identification of auditory channels that are dominated by the target signal (i.e., which portions of a sound mixture have a more favorable SNR) so that the output of these channels can be combined at a later stage to reconstruct the internal representation of that target. This view is largely consistent with the glimpsing model of speech recognition in noise (Cooke, 2003, 2006; Apoux and Healy, 2008, 2009, 2012; also see Miller and Licklider, 1950; Celmer and Bienvenue, 1987). In this model, speech recognition in noise relies on the ability to identify and group together time-frequency regions that contain a relatively undistorted view of local signal properties, the so-called glimpses. Consistent with the substantial effect of DC processing, it is suggested that this mechanism primarily relies on carrier disparities. These findings are also consistent with a dichotomy between the role of temporal envelope and TFS cues in speech recognition suggested previously by Smith *et al.* (2002) and Apoux *et al.* (2013). In this view, speech information is provided by the temporal envelope while the information

needed to extract the target signal from the sound mixture is primarily provided by the TFS. It should be noted, however, that other mechanisms are possibly involved in the segregation process as introducing carrier disparities was not sufficient to fully restore speech intelligibility in most conditions.

B. Contribution of resolved harmonics to the benefit from DC processing

One may argue that frequency components that are spectrally resolved by the peripheral auditory system contributed most to the benefit observed in the present study. Although spectrally resolved harmonics may have contributed to the segregation process, there is evidence that these cues are not required for the effects observed currently.

First, it should be noted that resolved harmonics typically refer to the ability to isolate the frequency components of a single harmonic complex tone. While the resolving power of the normal auditory system is still in debate, it is usually accepted that harmonics with numbers above eight are poorly, if at all, resolved (see [Moore and Gockel, 2011](#) for a review). However, in the present study, and for speech recognition in noise in general, the notion of spectrally resolved harmonics is not directly relevant. For concurrent speech signals, not only does the auditory system need to resolve the individual components for each of two harmonic complexes but it also needs to isolate the individual components of one signal from those of the other signal, hence necessitating a higher resolution than for harmonic complex tones presented in isolation. In fact, it may be assumed that the resolution needed to separate the individual components of two simultaneous harmonic complex tones is at least twice that needed for single harmonic complex tones. In situations involving two or more simultaneous talkers, it is probable that only a few, if any, harmonic components could be spectrally isolated. A mechanism based on spectrally resolved harmonics would therefore prove very ineffective. Yet NH listeners are usually able to maintain communication in such situations.

Second, the data in [Fig. 4](#) do not support an effect of resolved harmonics as typically assumed. If subjects were using spectrally resolved harmonics to segregate the two signals, intelligibility should have decreased significantly with increasing number of overlapping harmonics, producing clear dips in the functions. These dips should have been particularly visible in those conditions resulting in complete overlap of the target components (e.g., target/background combinations of 200/100, 300/100, or 300/150 pps). As confirmed by our statistical analysis, there was no significant dip in the functions shown in [Fig. 4](#). Moreover, a comparison between experiments 1 and 2 also shows that introducing harmonic relationships between the target and background carriers had no substantial effect on intelligibility as all the functions in experiment 2 were very similar to those in experiment 1 despite involving harmonically related rates. Taken together, the present data are largely inconsistent with a substantial contribution of spectrally resolved harmonics as the introduction of common multiples or divisors of the

target carrier rate did not lead to local variations in intelligibility.

This limited contribution and the low probability to resolve harmonics in the presence of competing talkers calls into question the general contribution of spectrally resolved harmonics to speech recognition in noise. One way to reconcile the present findings with the extensive literature on the role of harmonicity in segregation is to introduce a temporal factor in that harmonics could be isolated in the spectral or in the time domain. Such a mechanism is largely consistent with the glimpsing model. In this view, the same mechanism based on time-frequency glimpses could be employed in both the low- and high-frequency regions to segregate speech from noise. As a consequence, reduced frequency selectivity in CI users should not be a strong factor limiting benefit from DC processing as their temporal resolution is believed to be similar to, if not better than, NH listeners (e.g., [Shannon, 1992](#)).

Additional evidence comes from the data reported by [Deeks and Carlyon \(2004\)](#). In their study, the authors limited the frequency extent of their stimuli to the high-frequency region (>937 Hz) and used carriers with low F_0 s. This approach ensured that subjects would not be able to rely on spectrally resolved harmonics as the closely spaced frequency components are not resolved in this region. Still, [Deeks and Carlyon \(2004\)](#) observed improvements as large as 35% points, demonstrating that effective segregation can be achieved even in conditions for which listeners are likely unable to spectrally resolve the frequency components of the two signals.

Further evidence that spectrally resolved harmonics are not required for the observed benefit of DC processing is provided by the results of a pilot experiment assessing the effect of reduced frequency resolution. In this pilot experiment, CI users were presented with dual carrier stimuli. Due to their limited frequency resolution, it was difficult if not impossible for CI users to spectrally resolve the frequency components of the stimuli used in the present study even in the low-frequency region.⁷ In this experiment, direct audio input was delivered to the processor of three users of the Nucleus CP810 processor and ACE strategy. The number of channels, the number of maxima, and the stimulation rate (22 active electrodes, 10 maxima, 900 pps per channel) were then programmed using Nucleus Custom Sound software while leaving other parameters in their everyday settings. The method employed in experiments 1–3 was used to create the stimuli. Cochlear implant users were therefore presented with ten-band stimuli. However, each vocoder band was frequency aligned with one of the analysis channels of the implant by setting the center frequency of each vocoder band to match the center frequency of a corresponding implant channel as determined by the frequency-allocation table. The rate of the pulse train was 150 pps for the target and 100 pps for the background. These low rates were chosen to further limit resolvability. The lower stimulus rates were also preferable as the per-electrode stimulation rate was only 900 Hz for all subjects, therefore limiting the encoding of the stimuli. The reference SC condition consisted of the target and background envelopes imposed on a

single 150 pps pulse train. Prior to the experiment, each CI subject completed a practice session with stimuli processed in the SC condition. This session was divided into five blocks of ten sentences each with each block corresponding to a SNR from 3 to 15 dB in 3 dB steps. The practice blocks were completed in order of increasing difficulty. The SNR providing approximately 20% to 30% correct intelligibility was used for the actual experiment. Subjects then completed two experimental blocks in random order, one for SC and one for DC. Each block consisted of 25 AzBio sentences mixed with time-reversed sentences also from the AzBio set (different talkers). The first five mixtures from each block served as familiarization and were discarded. All three subjects showed a substantial benefit from DC processing. The individual benefits were 36%, 37%, and 24% points. Although a contribution from spectrally resolved harmonics cannot be completely excluded, the three sets of evidence discussed in the preceding text provide reasonable assurance that the improved segregation associated with DC processing does not require spectrally resolved harmonics and instead primarily reflects the ability of NH listeners and CI users to exploit timing differences between target and background pulses.

C. Implications for CI processing

As mentioned in Sec. I, a foreseeable and potentially promising implication for the DC approach involves implementation into CI speech processors where it may provide the TFS cues necessary for the user to actively extract speech from noise and improve speech recognition performance. The rationale for providing TFS cues to CI users is that these patients, in addition to their great difficulties understanding speech in noise, often show little benefit from fluctuations in the background noise (i.e., little masking release; Nelson *et al.*, 2003; Stickney *et al.*, 2004; Loizou *et al.*, 2009; Kwon *et al.*, 2012). Because this pattern is very similar to that found in NH studies examining recognition of speech with degraded TFS, it is believed that at least part of the difficulties that CI users have understanding speech in noise is due to the complete loss of TFS cues. Thus far, most effort in this regard has been directed toward encoding the original TFS of the acoustic signal to CI users. Encoding the original TFS, however, is technically challenging, and as a result, most approaches transmit only certain TFS-related cues (Riss *et al.*, 2008; Riss *et al.*, 2011; Li *et al.*, 2012), or they only present TFS acoustically via residual low-frequency hearing (electric-acoustic stimulation; Turner *et al.*, 2004; Dorman *et al.*, 2005; Kong *et al.*, 2005; Brown and Bacon, 2009a,b). Unfortunately, these TFS approaches have either resulted in limited improvements or can only be implemented in a small minority of patients. One clear advantage of the DC approach is that it does not require encoding of the original TFS. More importantly, the present study demonstrates how synthetic TFS information that is unrelated to the incoming sound mixture can be used to substantially improve speech intelligibility in noise.

Perhaps the most remarkable advantage of the DC approach is that it preserves the richness of the acoustic

environment while concurrently improving speech intelligibility in noise. Indeed the DC approach involves segregating the target signal from the background but preserving and transmitting both signals to the CI user. This is in stark contrast with most attempts to enhance speech intelligibility in noise in which the goal is generally to suppress the background. While the benefit of transmitting only the target signal is undeniable, CI users may value the increased awareness of the acoustic environment provided that intelligibility is satisfactory. Concomitantly, preserving the richness of the acoustic environment might also allow CI users to naturally direct their attention toward one of the signals in the environment. According to the present findings, an acceptable trade-off between intelligibility and awareness is within reach as the DC approach may allow CI users to achieve a level of speech understanding approaching or comparable to that of NH listeners (UNP, see Fig. 5) while preserving the complexity of the acoustic environment.

Finally, it is notable that the DC approach differs from most current strategies in that the goal is not to provide a clean target signal. The goal is to provide the CI user with a mixture of sound sources processed in a way that allows their auditory system to naturally extract the desired signal from the mixture, just as NH listeners are able to do. Previous approaches involving the extraction of target speech from the background and the suppression of any signal deemed less critical by the processor work under the assumption that the impaired auditory system is no longer capable of processing speech in noise effectively. Here we argue that although several major functions are severely compromised in the implanted auditory system, many computational functions remain intact. One reason why the implanted auditory system cannot take full advantage of its intact computational functions is because the signal delivered by the implant is severely impoverished. In other words, the auditory system of CI users may be perfectly able to separate speech from noise, it simply does not receive the cues necessary to do so. Therefore it may be assumed that by introducing in the encoded sound mixture the cues necessary to naturally extract the desired signal from the mixture, even the implanted auditory system could effectively process speech in noise.

ACKNOWLEDGMENTS

This research was supported by Grant No. DC008594 from the National Institute on Deafness and Other Communication Disorders. The authors are grateful to Christopher Brown for assistance collecting the CI data and to Brittney Carter for assistance collecting the NH data.

¹The isolated TFS was obtained by decomposing the signal in each of 16 frequency bands into envelope and TFS using the Hilbert transform and summing only the TFS components across bands.

²This was expected to be the case, just as simultaneous sounds are more difficult to separate if they share a same fundamental frequency (e.g., Broadbent and Ladefoged, 1957; Brokx and Nootboom, 1982; Scheffers, 1983; Assmann and Summerfield, 1990).

³The largest difference tested was that between 109 and 347 pps.

⁴The two values at approximately +10 and +20 pps above 197 were intended to be 211 and 223, respectively. The latter value, however, was accidentally set to 199, resulting in a 2 pps difference instead of 20.

- ⁵Deeks and Carlyon (2004) used complex harmonics with an F_0 of 40 and 70 Hz. By summing the harmonics in alternating phase, the waveform in each frequency channel resembled a modulated pulse train with a rate double the F_0 . Hence, the 80 and 140 Hz rates.
 - ⁶All methodological and procedural details were similar to those used in experiment 1 except that only five target rates (79, 113, 157, 179, and 197 pps) and two background rates (67 and 167 pps) were tested. For comprehensiveness, it should be noted that the four other target rates tested in this preliminary experiment (113, 157, 179, and 197 pps) yielded 8% to 43% points improvement over the single carrier condition.
 - ⁷This is especially true when considering that two harmonic complex tones were presented simultaneously. Moreover, the rate of these tones was 100 and 150 pps, resulting in a maximum frequency gap of 50 Hz.
- ANSI (2004). ANSI S3.21 (R2009), *American National Standard Methods for Manual Pure-Tone Threshold Audiometry* (Acoustical Society of America, New York).
- ANSI (2010). ANSI S3.6-2010, *American National Standard Specification for Audiometers* (Acoustical Society of America, New York).
- Apoux, F., and Bacon, S. P. (2008). "Differential contribution of envelope fluctuations across frequency to consonant identification in quiet," *J. Acoust. Soc. Am.* **123**, 2792–2800.
- Apoux, F., and Healy, E. W. (2008). "Phoneme recognition as a function of the number of auditory filter outputs," in *Proceedings of the Acoustics'08*.
- Apoux, F., and Healy, E. W. (2009). "On the number of auditory filter outputs needed to understand speech: Further evidence for auditory channel independence," *Hear. Res.* **255**, 99–108.
- Apoux, F., and Healy, E. W. (2010). "Relative contribution of off- and on-frequency spectral components of background noise to the masking of unprocessed and vocoded speech," *J. Acoust. Soc. Am.* **128**, 2075–2084.
- Apoux, F., and Healy, E. W. (2011). "Relative contribution of target and masker temporal fine structure to the unmasking of consonants in noise," *J. Acoust. Soc. Am.* **130**, 4044–4052.
- Apoux, F., and Healy, E. W. (2012). "Use of a compound approach to derive auditory-filter-wide frequency-importance functions for vowels and consonants," *J. Acoust. Soc. Am.* **132**, 1078–1087.
- Apoux, F., and Healy, E. W. (2013). "A glimpsing account of the role of temporal fine structure information in speech recognition," in *Basic Aspects of Hearing: Physiology and Perception*, edited by B. C. J. Moore, R. D. Patterson, I. M. Winters, R. P. Carlyon, and H. E. Gockel (Springer, New York), pp. 119–126.
- Apoux, F., Millman, R. E., Viemeister, N. F., Brown, C. A., and Bacon, S. P. (2011). "On the mechanisms involved in the recovery of envelope information from temporal fine structure," *J. Acoust. Soc. Am.* **130**, 273–282.
- Apoux, F., Yoho, S. E., Youngdahl, C. L., and Healy, E. W. (2013). "Role and relative contribution of envelope and temporal fine structure to sentence recognition in noise," *J. Acoust. Soc. Am.* **134**, 2205–2212.
- 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.
- Broadbent, D. E., and Ladefoged, P. (1957). "On the fusion of sounds reaching different sense organs," *J. Acoust. Soc. Am.* **29**, 708–710.
- Brokx, J. P. L., and Neeboom, S. G. (1982). "Intonation and the perception of simultaneous voices," *J. Phon.* **10**, 23–36.
- Brown, C. A., and Bacon, S. P. (2009a). "Low-frequency speech cues and simulated electric-acoustic hearing," *J. Acoust. Soc. Am.* **125**, 1658–1665.
- Brown, C. A., and Bacon, S. P. (2009b). "Achieving electric-acoustic benefit with a modulated tone," *Ear Hear.* **30**, 489–493.
- Celmer, R. D., and Bienvenue, G. R. (1987). "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Springer, The Netherlands), pp. 473–480.
- Cooke, M. (2003). "Glimpsing speech," *J. Phon.* **31**, 579–584.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.
- Deeks, J. M., and Carlyon, R. P. (2004). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," *J. Acoust. Soc. Am.* **115**, 1736–1746.
- Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (2005). "Acoustic simulations of combined electric and acoustic hearing (EAS)," *Ear Hear.* **26**, 371–380.
- Friesen, L. M., Shannon, R. V., Baskett, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Füllgrabe, C., Berthommier, F., and Lorenzi, C. (2006). "Masking release for consonant features in temporally fluctuating background noise," *Hear. Res.* **211**, 74–84.
- Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J. C. (2007). "Effect of spectral smearing on the perceptual segregation of vowel sequences," *Hear. Res.* **231**, 32–41.
- Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J. C. (2008). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," *J. Acoust. Soc. Am.* **124**, 3076–3087.
- Ghitza, O. (2001). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," *J. Acoust. Soc. Am.* **110**, 1628–1640.
- Gilbert, G., and Lorenzi, C. (2006). "The ability of listeners to use recovered envelope cues from speech fine structure," *J. Acoust. Soc. Am.* **119**, 2438–2444.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Gnansia, D., Pean, V., Meyer, B., and Lorenzi, C. (2009). "Effects of spectral smearing and temporal fine structure degradation on speech masking release," *J. Acoust. Soc. Am.* **125**, 4023–4033.
- Gnansia, D., Pressnitzer, D., Péan, V., Meyer, B., and Lorenzi, C. (2010). "Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees," *Hear. Res.* **265**, 46–53.
- Healy, E. W., and Steinbach, H. M. (2007). "The effect of smoothing filter slope and spectral frequency on temporal speech information," *J. Acoust. Soc. Am.* **121**, 1177–1181.
- Hopkins, K., and Moore, B. C. J. (2010). "The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.* **127**, 1595–1608.
- IEEE (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Kong, Y. Y., Stickney, G. S., and Zeng, F.-G. (2005). "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **117**, 1351–1361.
- Kwon, B. J., Perry, T. T., Wilhelm, C. L., and Healy, E. W. (2012). "Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear implant listeners," *J. Acoust. Soc. Am.* **131**, 3111–3119.
- Li, X., Nie, K., Imennov, N. S., Won, J. H., Drenna, W. R., Rubinstein, J. T., and Atlas, L. E. (2012). "Improved perception of speech in noise and Mandarin tones with acoustic simulations of harmonic coding for cochlear implants," *J. Acoust. Soc. Am.* **132**, 3387–3398.
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. (2009). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting," *J. Acoust. Soc. Am.* **125**, 372–383.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18866–18869.
- Miller, G. A., and Licklider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Mitterbacher, A., Zierhofer, C., Schatzer, R., and Kals, M. (2005a). "Encoding fine time structure with channel specific sampling sequences," in *Conference on Implantable Auditory Prosthesis*, Pacific Grove, CA.
- Mitterbacher, A., Zierhofer, C., Schatzer, R., Kals, M., Nopp, P., Schleich, P., Krebelder, U., and Nobbe, A. (2005b). "Pitch, fine structure and CSSS—Results from patient tests," in *British Cochlear Implant Group Academic Meeting*, Birmingham, UK.
- Moore, B. C., Glasberg, B. R., and Peters, R. W. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* **80**, 479–483.
- Moore, B. C., and Gockel, H. E. (2011). "Resolvability of components in complex tones and implications for theories of pitch perception," *Hear. Res.* **276**, 88–97.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.

- Qin, M. K., and Oxenham, A. J. (2006). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," *J. Acoust. Soc. Am.* **119**, 2417–2426.
- Riss, D., Arnoldner, C., Baumgartner, W. D., Kaider, A., and Hamzavi, J. S. (2008). "A new fine structure speech coding strategy: Speech perception at a reduced number of channels," *Otol. Neurotol.* **29**, 784–788.
- Riss, D., Hamzavi, J. S., Selberherr, A., Kaider, A., Blineder, M., Starlinger, V., Gstottner, W., and Arnoldner, C. (2011). "Envelope versus fine structure speech coding strategy: A crossover study," *Otol. Neurotol.* **32**, 1094–1101.
- Scheffers, M. T. M. (1983). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.
- Shannon, R. V. (1992). "Temporal modulation transfer functions in patients with cochlear implants," *J. Acoust. Soc. Am.* **91**, 2156–2164.
- Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). "Chimaeric sounds reveal dichotomies in auditory perception," *Nature* **416**, 87–90.
- Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., and Cook, S. (2012). "Development and validation of the AzBio sentence lists," *Ear Hear.* **33**, 112–117.
- Stickney, G. S., Nie, K., and Zeng, F.-G. (2005). "Contribution of frequency modulation to speech recognition in noise," *J. Acoust. Soc. Am.* **118**, 2412–2420.
- Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Studebaker, G. A. (1985). "A rationalized arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Swaminathan, J., Reed, C. M., Desloge, J. G., Braida, L. D., and Delhorne, L. A. (2014). "Consonant identification using temporal fine structure and recovered envelope cues," *J. Acoust. Soc. Am.* **135**, 2078–2090.
- Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (2004). "Speech recognition in noise for cochlear-implant listeners: Benefits of residual acoustic hearing," *J. Acoust. Soc. Am.* **115**, 1729–1735.
- Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H. (2004). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," *J. Acoust. Soc. Am.* **116**, 1351–1354.