A simulation study of harmonics regeneration in noise reduction for electric and acoustic stimulation

Yi Hu^{a)}

Department of Electrical Engineering, University of Texas–Dallas, Richardson, Texas 75080

(Received 7 May 2009; revised 25 February 2010; accepted 2 March 2010)

Recent research results show that combined electric and acoustic stimulation (EAS) significantly improves speech recognition in noise, and it is generally established that access to the improved F0 representation of target speech, along with the glimpse cues, provide the EAS benefits. Under noisy listening conditions, noise signals degrade these important cues by introducing undesired temporal-frequency components and corrupting harmonics structure. In this study, the potential of combining noise reduction and harmonics regeneration techniques was investigated to further improve speech intelligibility in noise by providing improved beneficial cues for EAS. Three hypotheses were tested: (1) noise reduction methods can improve speech intelligibility in noise for EAS; (2) harmonics regeneration after noise reduction can further improve speech intelligibility in noise for EAS; and (3) harmonics sideband constraints in frequency domain (or equivalently, amplitude modulation in temporal domain), even deterministic ones, can provide additional benefits. Test results demonstrate that combining noise reduction and harmonics regeneration can significantly improve speech recognition in noise for EAS, and it is also beneficial to preserve the harmonics sidebands under adverse listening conditions. This finding warrants further work into the development of algorithms that regenerate harmonics and the related sidebands for EAS processing under noisy conditions. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3372718]

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

Pages: 3145–3153

I. INTRODUCTION

In the past decade there has been significant progress in the technology of combining hearing aids and cochlear implants to provide further benefits to cochlear implant users who have preserved residual acoustic hearing in the low frequency. For these patients, access to low-frequency information is provided by hearing aids using acoustic stimulation; high-frequency information is presented by cochlear implants using electric stimulation.

Numerous studies of combining electric and acoustic stimulation (EAS) have demonstrated the significant benefits facilitated by acoustic stimulation (vonIlberg *et al.*, 1999; Turner *et al.*, 2004; Kiefer *et al.*, 2005; Kong *et al.*, 2005; Gantz *et al.*, 2005, 2006; Dorman *et al.*, 2008), all of which involved EAS patients. In a recent multicenter clinical trial, Gantz *et al.* (2006) showed that EAS patients fitted with Hybrid short-electrode cochlear implant gained improved word understanding in noise. When compared with a group of high-performance, long-electrode implant users who were matched in terms of word recognition in quiet, Hybrid users achieved a mean advantage of 6.1 dB in multitalker babble and 7.5 dB in steady-state noise. Kong *et al.* (2005) tested bimodal EAS users wearing a hearing aid in one ear and a cochlear implant in the other. These patients also showed significant improvements in both melody recognition and speech recognition in competing talker background. More

recently, Dorman *et al.* (2008) conducted a comprehensive study on speech, voice, and melody recognition by bimodal EAS users. These bimodal users also demonstrated significant improvements in terms of melody recognition and speech recognition both in quiet and in noise, although the voice discrimination results were inconclusive.

To study the contributions of low-frequency acoustic information, simulated EAS experiments with normal-hearing subjects have been introduced to examine the mechanisms underlying the EAS benefits using controlled simulation conditions (Dorman *et al.*, 2005; Chang *et al.*, 2006; Qin and Oxenham, 2006; Kong and Carlyon, 2007; Li and Loizou, 2008; Brown and Bacon, 2009). As expected, all these studies reported synergistic EAS benefits for speech recognition in noise; more importantly, they also proposed several hypotheses about the beneficial low-frequency cues and the underlying mechanisms. Chang *et al.* (2006) hypothesized that F0 present in low-frequency sound helped to group the speech information from high-frequency temporal envelopes. Qin and Oxenham (2006) also suggested that the EAS benefits were in part attributed to a better F0 representation. The study by Kong and Carlyon (2007) presented evidence that F0 was neither sufficient nor necessary for the observed EAS benefits, although some of the results were obtained in conditions with noise absent in low frequency. Kong and Carlyon concluded that EAS benefits were derived from voicing and glimpsing cues. Li and Loizou (2008) presented the hypothesis that a glimpsing mechanism was responsible for the observed EAS benefits. Most recently, Brown and Bacon (2009) presented their results that confirmed the F0 importance of the target speech for EAS and drew conclusions that supported a glimpsing account of EAS.

^{a)}Author to whom correspondence should be addressed. Present address: Department of Electrical Engineering and Computer Science, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201. Electronic mail: huy@uwm.edu

These results so far demonstrate that better glimpsing and F0 representation from low-frequency sound are two important factors in EAS benefits; however, one often ignored issue is that although F0 and glimpsing are robust to noise, they also degrade in noise. Noise degrades the harmonic structure in general [Figs. 6 and 10 in [Zavarehei et al., \(2007\)](#)] and consequently degrades F0 representation. Also, when a noisy situation becomes worse, glimpsing becomes more difficult ([Cooke, 2006](#), Fig. 1).

The motivation of our present study is that if glimpsing and F0 representation from noisy speech are improved, further EAS benefits can be realized. To this end, we investigated two techniques, noise reduction and harmonics regeneration, which were used to target the two established mechanisms underlying EAS benefits. More specifically, noise reduction was used to improve glimpsing and harmonics regeneration was used to improve both F0 representation and glimpsing in voiced segments.

The glimpsing model of speech perception in noise proposed by [Cooke \(2006\)](#) is often associated with the work on ideal binary masking ([Wang, 2005](#); [Brungart et al., 2006](#); [Li and Loizou, 2007](#)), which clearly improves glimpsing when used as a noise reduction method under noisy conditions. In practical applications, estimates of binary masks rely on conventional noise reduction methods, which do not perform adequately for this purpose (Table III in [Hu and Loizou, 2008](#)). For the noise reduction method in the present study, we adopted a soft mask approach that applies continuous gain function values ranging between 0 and 1, which are in contrast to the binary gain values in the ideal binary mask approach, and work better with conventional noise reduction methods. After noise reduction, glimpsing cues improve in the low-frequency region because reduced noise level makes glimpsing easier. In the high-frequency region where speech information is mainly vocoded in the temporal envelopes, it has been shown in many studies that noise reduction methods improve temporal envelopes for vocoded speech ([Yang and Fu, 2005](#); [Loizou et al., 2005](#); [Hu et al., 2007](#)).

Noise reduction methods improve glimpsing cues in the low-frequency region; however, most noise reduction algorithms are based on preset optimization criteria and treat all frequency components equally ([Loizou, 2007](#)). In other words, preservation of harmonic structures and the associated sidebands (see more details in Sec. II A 3) is not taken into account during noise reduction process. Consequently, in most of the cases, harmonics and the associated sidebands further degrade after noise reduction ([Plapous et al., 2006](#); [Zavarehei et al., 2007](#)). We hypothesize that a deliberately designed post-processing step that regenerates harmonics in the acoustic portion can improve F0 representation, and probably facilitate glimpsing in the voiced region as well. In addition, the sidebands associated with the clean harmonics are of similar shape, which can be viewed as a constraint. This constraint can be properly modeled and restored after noise reduction. We hypothesize that the sideband restoration after harmonics regeneration can further improve EAS benefits.

In summary, the present study tests three hypotheses: (1) noise reduction methods can improve speech intelligibility in

FIG. 1. (Color online) Histogram of the estimated F0 values in 10 ms frames.

noise for EAS; (2) harmonics regeneration after noise reduction can further improve speech intelligibility in noise for EAS; and (3) restoration of harmonics sideband constraints can provide additional EAS benefits.

II. EXPERIMENT: NOISE REDUCTION AND HARMONICS REGENERATION

A. Methods

1. Subjects

Nine normal-hearing native speakers of American English (five males, four females) participated in this experiment. All subjects were paid an hourly wage for their participation. Subjects' age ranged from 20 to 36 years, and all of them were undergraduate and graduate students at the UT-Dallas campus.

2. Stimuli

The target speech materials consisted of sentences from the IEEE database ([IEEE, 1969](#)) and were obtained from [Loizou \(2007\)](#). The IEEE corpus has 72 lists of ten phonetically balanced sentences of high difficulty. The sentences were produced by a male speaker and recorded in a double-walled sound-attenuation booth at a sampling rate of 25 kHz. Figure 1 shows the histogram of the estimated F0 values in 10 ms frames for this male speaker (see Sec. II A 3 for more details about F0 detection). The estimated F0 values ranged from 75 to 250 Hz, with the mean equal to 127.57 Hz and the standard deviation equal to 21.16 Hz.

The masker used was speech-shaped noise, a type of continuous (steady-state) noise that had the same long-term spectrum as the sentences in the IEEE corpus. Speech-shaped noise was selected because it has been shown to be a very effective masker ([Turner et al., 2004](#); [Li and Loizou, 2008](#); [Brown and Bacon, 2009](#)), and its stationarity minimizes the confounding effect of the accuracy of noise estimation algorithms (see Sec. II A 3 for more details about noise estimation).

TABLE I. Cutoff frequencies for the channel allocation.

Channel	Unprocessed (below 600 Hz)	
	Low (Hz)	High (Hz)
1	724	1158
2	1158	1790
3	1790	2710
4	2710	4050
5	4050	6000

3. Signal processing

The experiments were designed to evaluate the benefits of noise reduction and harmonics regeneration for EAS in noise; five processing conditions were used for this purpose. The first processing condition was designed to simulate the combined electric and acoustic stimulation. As the first step, a pre-emphasis filter with 2000 Hz cutoff and 3 dB/octave rolloff was applied to the signal. For the acoustic stimulation, the speech signal was lowpass filtered up to 600 Hz using a sixth-order Butterworth filter. For the electric stimulation above 600 Hz, a five-channel noise-excited vocoder was used (Shannon *et al.*, 1995); the speech signal was band-passed into five frequency bands between 724 Hz and 6000 Hz using sixth-order Butterworth filters. For the specified frequency range, the Cambridge filter spacing (Glasberg and Moore, 1990) was used to allocate the five frequency channels (see Table I). The envelopes of the bandpassed signals were obtained by full-wave rectification, followed by low-pass filtering using a second-order Butterworth filter with a 400 Hz cutoff frequency. F0 cues in the temporal envelopes were preserved with this 400 Hz cutoff frequency (see Fig. 1). The independent white-noise carriers were modulated by the extracted temporal envelopes. The same bandpass filters (as in Table I) were then used to filter the amplitude-modulated white noises. The narrow-band modulated noises were finally summed and combined with the lowpass filtered speech signal to generate the EAS stimuli. The level of the synthesized speech signal was scaled to have the same root mean square value as the original speech signal.

The other four conditions involved a pre-processing step prior to EAS processing. The second processing condition was designed to test the hypothesis that a noise reduction method can benefit EAS in noise. A noise reduction method based on minimum mean-square error log-spectral amplitude estimation (LogMMSE) was used in the pre-processing step. LogMMSE proposed by Ephraim and Malah (1985) was selected because it has performed well in both speech quality and speech intelligibility tests (Hu and Loizou, 2007a, 2007b). Noise estimation that monitors the constantly changing noise spectrum values is an essential component in noise reduction, and it can be a very challenging task in a low-signal-to-noise-ratio (SNR), non-stationary noise environment. In the evaluation of noise reduction methods, the accuracy of noise estimation is a confounding factor since suppression function values applied to noisy signal depend on estimated noise spectrum; for this reason, noise estima-

FIG. 2. (Color online) Block diagram for LogMMSE.

tion is still a very active research topic. In this study, the same noise estimation method as in Hu and Loizou (2007a, 2007b) was used. Short-time Fourier transform-based frame processing was used in LogMMSE. Speech signals were segmented into 50% overlapping frames by being multiplied with a sliding 20-ms Hanning window. Figure 2 shows the block diagram for the frame processing in LogMMSE. A 8192-point FFT (by zero padding) with a frequency bin resolution of 3.05 Hz was used.

The third processing condition was designed to evaluate the benefit of harmonic regeneration when combined with noise reduction. Our hypothesis is that by regenerating harmonics after noise reduction, speech intelligibility can be further improved for EAS in noise. The majority of speech signals are voiced segments (Mines *et al.*, 1978), which can be approximately modeled by harmonics. In a typical voiced frame, speech sound is produced by inputting a periodic glottal source waveform into the vocal tract that determines the formants (F1/F2/F3), and the harmonics appear at the integer multiples of F0. Under a noisy environment, the noise signal degrades the harmonic structure. Noise reduction methods further degrade the harmonic structure due to the inaccurate noise spectrum estimation. Since voiced frames have much higher local SNRs compared with unvoiced frames, the estimation of voiced cues is generally much more robust. Many robust F0 detection algorithms exist and a harmonic regeneration operator can be based on adaptive comb filtering (Nehorai and Porat, 1986), nonlinear functions (Plapous *et al.*, 2006) and codebook approach (because harmonics amplitudes are highly correlated; see more details in Chu, 2004; Zavarehei *et al.*, 2007, Fig. 2).

In our present study, to establish an upper bound and evaluate the potential of the harmonics regeneration when combined with noise reduction, we assume an ideal operating environment, i.e., we estimate F0 from clean speech and regenerate harmonics with prior knowledge from clean speech spectrum. More specifically, an F0 detection algorithm based on autocorrelation function (Kondoz, 1999) was used to obtain F0. The number p of harmonics regenerated was then calculated by $p = \lfloor 600/F0 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor operator. To compensate for the possible inaccuracy of pitch

FIG. 3. (Color online) Block diagram for the combined harmonics regeneration and noise reduction.

detector, harmonics were regenerated by extracting the local peaks in a 30-Hz range around nF_0 , where $n=1, \dots, p$, and the extracted harmonics have a quantization error of roughly 1.53 Hz (half of the FFT frequency resolution). Figure 3 shows the block diagram for the combined harmonics regeneration and noise reduction.

The fourth processing condition was designed to further enhance harmonics regeneration. An inspection of the clean harmonics in frequency domain reveals an important observation that they share a similar sideband shape, and it is expected that both noise signal and noise reduction degrade this sideband constraint. We hypothesize that this sideband constraint can be exploited as a beneficial cue. There are two ways to model the sideband constraint: one approach is to assume the sideband shape depends on the harmonic amplitudes (Zavarehei *et al.*, 2007); the other approach is to assume the sideband shape is deterministic. Although the first approach is more accurate and may lead to better performance, we selected the second approach in the present study because a deterministic sideband constraint had an intuitive interpretation and led to important implications. More specifically, the harmonic sidebands were generated by multiplying each one of the harmonics with the Gaussian-shaped function, which spans roughly 75 Hz (this Gaussian-shaped function was determined by visually inspecting the harmonics in several voiced frames). Figure 4 illustrates the process to regenerate harmonics with sideband. The process is equivalent to a convolution in the frequency domain between the band-limited, Gaussian-shaped function and the regenerated harmonics. As the convolution operation in frequency domain is equivalent to the modulation operation in time domain (Mitra, 2001, Table III.5), this process is also equivalent to the multiplication of two terms in time domain: one is the inverse Fourier Transform of the band-limited, Gaussian-shaped function, which can be viewed as an amplitude

FIG. 4. The process to regenerate harmonics with sideband, which is equal to a convolution operation between the regenerated harmonics and the Gaussian function. The top panel shows the regenerated harmonics. The middle panel shows the band-limited, Gaussian-shaped function. The bottom panel shows regenerated harmonics with sideband.

modulation function or a temporal envelope; the other is the inverse Fourier transform of the regenerated harmonics, which can be viewed as a temporal fine structure.

The fifth processing condition was designed to measure the intelligibility of the regenerated harmonics that carry sidebands. In voiced frames, harmonics with sidebands were regenerated as in conditions three and four. It is expected that no speech intelligibility exists in the regenerated harmonics carrying sidebands.

4. Procedure

The subjects were seated in a double-walled sound-attenuation booth (Acoustic Systems, Inc.). The tests were conducted using a PC connected to a Tucker-Davis system 3. Stimuli were played monaurally to the subjects through Sennheiser HD 250 Linear II circumaural headphones at a comfortable listening level. To familiarize each subject with the stimuli, a training session was administered prior to the formal testing and each subject listened to EAS-processed speech stimuli. The training session typically lasted about 15 to 20 min. During the test, the subjects were instructed to write down the words they heard. In total, there were 17 testing conditions ($=4$ SNR levels $\times 4$ processing methods $+1$ condition for clean harmonics with sidebands). For each condition, two lists of sentences were used, none of which were repeated across the testing conditions. The conditions were presented in random order for each subject. Between conditions, subjects were allowed to take breaks at their leisure and no feedback was provided after each testing condition.

B. Results

The mean percent correct scores for all conditions are shown in Fig. 5, where Noproc is the first processing condition (Sec. II A 3) to simulate the combined electric and

FIG. 5. (Color online) Mean percent correct scores as a function of SNR level for speech-shaped noise. The error bars denote ± 1 standard error of the mean.

acoustic stimulation. LogMMSE is the second condition to test the hypothesis that noise reduction can benefit EAS in noise. LogMMSE+Har0 is the third condition, which evaluates the benefit of harmonic regeneration when combined with noise reduction. LogMMSE+Har3 is the fourth condition, which regenerates the clean harmonics with sideband. The score for the fifth condition (Har3) to use only clean harmonics with sidebands was replicated at each SNR level to make a better contrast. It can be seen that the clean harmonics itself does not carry any speech intelligibility. Performance was measured in terms of the percent of words identified correctly (all words were scored except the/a/an). Statistical significance was determined by using percent correct score as the dependent variable, and SNR levels and processing conditions as the two within-subjects factors. Two-way analysis of variance with repeated measures indicated significant effects of both SNR levels [$F(3,24) = 142, p < 0.001$] and processing conditions [$F(3,24) = 211.55, p < 0.001$], and significant interaction between SNR levels and processing conditions [$F(9,72) = 4.61, p < 0.001$].

Multiple paired comparisons with Bonferroni correction were run between the scores obtained with Noproc and LogMMSE, LogMMSE and LogMMSE+Har3, LogMMSE and LogMMSE+Har0, and LogMMSE+Har3 and LogMMSE+Har0 at the various SNR levels. The Bonferroni-corrected statistical significance level was set at $p < 0.0125$ ($\alpha = 0.05$). The comparison results are shown in Table II. The comparisons indicated statistically significant differences between the Noproc and LogMMSE scores at any of the four SNR levels ($p < 0.007$), suggesting that noise reduction methods can benefit EAS in steady-state noise. The

TABLE II. Multiple paired comparison results at the various SNR levels.

	-2 dB	0 dB	2 dB	4 dB
LogMMSE—Noproc	**a	**	**	**
LogMMSE+Har3—LogMMSE	**	**	**	**
LogMMSE+Har0—LogMMSE	**	**		
LogMMSE+Har3—LogMMSE+Har0	**			**

a** indicates significantly better ($p < 0.0125$).

scores obtained with the LogMMSE+Har0 stimuli at lower SNR levels (-2 and 0 dB) were significantly higher ($p < 0.003$) than those obtained with the LogMMSE stimuli, but they did not differ ($p > 0.75$) at higher SNR levels (2 and 4 dB), suggesting that introducing harmonics structure can further improve the benefits with noise reduction methods at lower SNR levels. The scores obtained with the LogMMSE+Har3 stimuli were significantly higher than those obtained with the LogMMSE stimuli at SNR levels of -2, 0, and 4 dB ($p < 0.006$), and marginally higher at 2 dB ($p = 0.01$). The scores obtained with the LogMMSE+Har3 stimuli at SNR levels of -2 and 4 dB were significantly higher ($p < 0.005$) than those obtained with the LogMMSE+Har0 stimuli, but they did not differ ($p > 0.2$) at SNR levels of 0 and 2 dB.

The comparison results indicate that noise reduction is a viable method to improve EAS under noisy listening conditions. Although the harmonics structure itself does not carry any speech intelligibility, its existence can further benefit noise reduction methods. These results suggest that combining harmonics regeneration and noise reduction has the potential to significantly improve EAS in noise.

III. GENERAL DISCUSSION AND CONCLUSIONS

The present experiment results indicate that the LogMMSE noise reduction method provided significant speech intelligibility improvement for EAS listening in steady-state noise. A mean advantage of almost 17 percentage points were obtained over the unprocessed condition across the four SNR levels. This outcome is most likely attributed to the combined effects of improved glimpsing in the low-frequency acoustic portion (Li and Loizou, 2008) and enhanced temporal envelope cues in the high-frequency region (Yang and Fu, 2005; Loizou et al., 2005; Hu et al., 2007). On the other hand, noise reduction algorithms degrade harmonic structures, and, as a result, make F0 representation poorer. This is especially true at lower SNR levels, as the impact of inaccurate noise estimation is more significant. To overcome this disadvantage, harmonics regeneration can be integrated into noise reduction to improve F0 representation. From the experiment results, harmonics regeneration yields an additional mean advantage of almost 12 percentage points over the noise reduction method at the two lower SNR levels, although no benefits were observed at the two higher SNR levels as more accurate noise estimation did not degrade the harmonic structure. As explained in Sec. II A 3, harmonics are usually associated with similarly shaped sidebands, which are vulnerable to noise reduction methods due to inaccurate noise estimation. By integrating the sideband constraints into harmonics regeneration, we achieved an advantage of 28 percentage points at -2 dB, 12 percentage points at 0 dB, 6 percentage points at 2 dB, and 15 percentage points at 4 dB, compared with the LogMMSE noise reduction method. Figure 3 shows that, after LogMMSE noise reduction, if the current frame is voiced, the clean harmonics below 600 Hz are regenerated and the other enhanced frequency components remain intact. In other words, the frequency components above 600 Hz in voiced frames and all the frequency components in unvoiced frames

are still from LogMMSE noise reduction method; thus there might be an interaction between the effects of the clean harmonics regeneration algorithm and the LogMMSE noise reduction method. The wide performance range observed at different SNR levels was most likely caused by this interaction.

The results of the simulation experiments described above demonstrate that combining harmonics regeneration and noise reduction can achieve a mean advantage of more than 32 percentage points across the four tested SNR levels over the unprocessed EAS condition, and it is a viable approach to improve speech intelligibility for EAS listening in noise. It should be pointed out that our simulations assume “flat” frequency response in the low-frequency acoustic region, which is clearly not true for EAS patients who generally have a sloping frequency response in low frequency. Also, the perception of pitch is different between hearing-impaired people and normal-hearing people (Moore and Carlyon, 2005); for instance, reduced frequency selectivity was found in listeners with cochlear hearing loss, and according to the place theory, this might have a negative impact on the capability to resolve partials in harmonics complex and perceive pitch; in addition if dead regions exist within the cochlear and some frequency components fall into the dead regions, they may be detected at places quite different from the places in normal cochlear; therefore, the pitch perception can be quite different. On the other hand, acoustic simulations have shown predictive capabilities when it comes to interpreting EAS benefits (Turner *et al.*, 2004; Dorman *et al.*, 2005; Kong *et al.*, 2005). With this in mind, the following section presents more details about the improvements of glimpsing by noise reduction, the importance of harmonics regeneration, several practical considerations in algorithm implementation, and the implications for future EAS processor design.

A. Glimpsing

The glimpsing model of speech perception in noise by Cooke (2006) was based on two characteristics of speech signal: sparseness, which refers to the fact that high-energy regions are sparsely distributed; and redundancy, which refers to the fact that speech intelligibility is largely intact after information decimation. Glimpsing involves two steps: detection of the target, followed by integration of the cues from the glimpses. Detecting the target is accomplished by comparing the local SNR values of the time-frequency regions to a preset threshold. Local SNR is defined as the ratio of the speech energy to the noise energy. In the second step, the glimpses from the detection process are patched together using multiple cues such as F0/F1/F2 information. It remains unclear what a “useful” glimpse is in terms of duration and time-frequency locations.

A glimpsing mechanism was suggested to underlie the benefits of EAS (Kong and Carlyon, 2007; Li and Loizou, 2008; Brown and Bacon, 2009). Li and Loizou (2008) argued that the key contributing factor was the low-frequency SNR advantage inherent in the speech signal, especially for voiced speech segments. This SNR advantage and better

spectral resolution (more glimpse patches) make glimpsing an easier task in the low-frequency acoustic region. It can be seen from Li and Loizou (2008, Fig. 3) that this SNR advantage is much more muted for speech-shaped noise compared with a female talker, which explains significantly better EAS listening in competing talker background than in speech-shaped noise.

This is not surprising considering that competing talker background has a sparse distribution of high-energy regions; therefore, it only masks the target speech signal in a small portion of the time-frequency regions, while speech-shaped noise does not and it is a more effective masker.

It is well known that noise reduction algorithms improve SNRs by reducing noise levels. Therefore, after noise reduction, when compared with Li and Loizou (2008, Fig. 3), an elevated SNR profile is obtained for speech-shaped noise in both the low-frequency acoustic portion and the high-frequency electric region. In the low-frequency acoustic portion, compared with unprocessed EAS stimuli, the LogMMSE noise reduction method provides an incremental SNR advantage, which makes glimpsing easier and results in more glimpse patches. Meanwhile, in the high-frequency electric portion, the additional SNR advantage improves temporal envelopes (Hu *et al.*, 2007), which integrate with the enhanced glimpsing in low frequency to provide improved EAS benefits.

B. Harmonics regeneration

The majority of speech signals are voiced segments, which are composed of mainly harmonics and provide prosodic cues. Under noisy conditions, voiced segments enjoy much higher SNRs compared with unvoiced segments; glimpsing is also easier in voiced segments. From a speech synthesis point of view, harmonics are easy to generate because the only essential components are F0 and formant structures (F1/F2/F3), and many algorithms exist to estimate both (Kondoz, 1999; Huang *et al.*, 2001). This makes the harmonic plus noise model (HNM) a popular choice for a speech synthesis system (Stylianou, 2001).

Harmonics are important in pitch perception and good F0 discrimination requires the presence of low-numbered resolved harmonics (Plack and Oxenham, 2005). In EAS hearing, the low-frequency acoustic portion provides access to several low-numbered harmonics (the harmonic number depends on the F0 value), and this harmonic structure consists of F0 and F1 information. Qin and Oxenham (2006) showed that at a cutoff frequency of 300 Hz, it was reasonable to conclude that an improvement in F0 representation aided EAS performance; at a cutoff frequency of 600 Hz, part of the EAS benefits was likely due to improved spectral representation of F1. Hence, preservation of the harmonic structure is critical to F0 representation and likely benefits F1 representation.

Although noise reduction improves glimpsing, harmonic structures degrade to some degree in the process because noise reduction in general does not use any optimization criteria that preserve harmonic structures. To remedy this matter, a post-processing harmonic regeneration step is applied

after noise reduction (Nehorai and Porat, 1986; Plapous *et al.*, 2006; Zavarehei *et al.*, 2007). In frequency domain, from a signal approximation perspective, harmonics regeneration combined with noise reduction align the processed signals better with the speech production model; while in temporal domain, temporal fine structures that carry periodicity information are partly restored. The accompanying benefits are threefold: first, F0 representation is improved; second, F1 representation is likely better; third, glimpsing improves in the low-frequency voiced segments.

All the harmonics carry sidebands. From a signal processing perspective, in the frequency domain, the harmonics and the associated sidebands can be generated by multiplying the harmonics with band-limited functions; in the temporal domain, the harmonics are temporal fine structures and the sidebands are temporal envelopes. In HNM (Zavarehei *et al.*, 2007; Stylianou, 2001), these band-limited functions are modeled by a weighted sum of a Gaussian-shaped function and Rayleigh distribution, and the weight is determined by the harmonicity. In the present study, the band-limited function is Gaussian-shaped and all the harmonics use the same band-limited function to generate the sidebands. This arrangement yields a very simple sideband constraint that applies a deterministic temporal envelope to the temporal fine structures representing the harmonics. The experiment results demonstrate that this intuitive sideband constraint can serve as a beneficial cue for EAS hearing in noise.

C. Practical considerations in algorithm implementation

The present study demonstrated the full potential of regenerating speech signals' harmonic structure prior to EAS processing by using the harmonics from clean speech signals. The proposed algorithm involves two steps. The first step is the implementation of a noise suppression algorithm LogMMSE, which can be easily integrated into existing speech coding strategies such as ACE that use FFT processing to extract channel envelopes. Although many studies on ideal binary masking (Wang, 2005; Brungart *et al.*, 2006; Li and Loizou, 2007; Hu and Loizou, 2008) demonstrated restoration of speech intelligibility in noise, they assumed access to both clean and noise signals. Without this assumption, the deployment of ideal binary masking techniques demands substantial computation resources that involve machine learning techniques as reported in some studies (Kim *et al.*, 2009; Hu and Loizou, 2009), otherwise the conventional noise estimation techniques have not been shown to be effective in estimating binary masks (Hu and Loizou, 2008). For this reason, a soft mask approach of low computational cost was selected in the present study.

The second step of the proposed algorithm is regenerating the harmonic structure in the voiced segments of the enhanced speech signals. For this purpose, we assumed that we had access to F0 values and harmonic amplitudes extracted from clean speech. In practical applications, the F0 has to be estimated in real time; recently, Zakis *et al.* (2007) proposed a real-time F0 estimator suitable for CI processing. Reasonably accurate F0 detectors exist at the SNR levels in most practical scenarios; for instance, the F0 detection algo-

rithm in Zavarehei *et al.* (2007) produced 17%, 11%, and 6% estimation errors in 0, 5, and 10 dB train noise, and 21%, 13%, and 7% errors in 0, 5, and 10 dB babble noise. Although 21% error (in 0 dB SNR babble noise) in F0 estimation could cause problem in harmonics regeneration, considering most real-world applications operate at SNR levels of 5 dB and above, the state-of-art F0 detectors are reasonably accurate, as have been demonstrated in various real-world cellular phone applications (Kondoz, 1999); in addition, machine learning techniques used in Hu and Loizou (2009) can be beneficial to obtain better F0 estimation. The impact of the F0 estimation errors on speech quality and intelligibility for EAS processing warrants further investigation.

Several methods exist to estimate the harmonic amplitudes from a noisy signal. One simple approach is to estimate the formant structure (Huang *et al.*, 2001), followed by a rule-based formant synthesizer proposed by Klatt (1980). Alternatively, harmonic amplitudes can be estimated using codebook-based techniques (Chu, 2004; Zavarehei *et al.*, 2007), nonlinear functions (Plapous *et al.*, 2006), and adaptive comb filtering techniques (Nehorai and Porat, 1986). After the harmonic amplitudes are estimated, generating the harmonic sidebands, as shown in Fig. 4, is straightforward.

It should be pointed out that in the present study only combined electric and acoustic stimulation benefits were measured, and no separate measurements were made for the electric stimulation or acoustic stimulation conditions; thus there is no direct evidence to show whether the benefits come from electric stimulation or acoustic stimulation or both. However, past studies (Yang and Fu, 2005; Hu *et al.*, 2007) showed that noise reduction benefits electric stimulation; and it is most probable that noise reduction improves glimpsing in acoustic stimulation, so it is reasonable to expect that in the first step of noise reduction, the benefits come from both electric stimulation and acoustic stimulation. In the second step, the harmonic regeneration is only conducted in acoustic stimulation, therefore the extra benefits come from acoustic stimulation only.

In the present study, only a portion of the harmonic structure in the voiced segments were actually regenerated, depending on the F0 values. To keep the processing uniform, the number of regenerated harmonics was obtained by flooring the ratio between 600 Hz and the F0 value; the Gaussian-shaped function also was band-limited to 75 Hz, as the estimated pitch values ranged from 75 Hz to 250 Hz. This processing might be too conservative for some voiced segments, and a viable alternative is to synthesize all the frequency components below 600 Hz or the whole voiced segments using F0 values, harmonic amplitudes, and sidebands information.

D. Implications for EAS in noise

There are several implications of our study for EAS in noise. As in traditional cochlear implants (Yang and Fu, 2005; Loizou *et al.*, 2005; Hu *et al.*, 2007), noise reduction can improve speech intelligibility in steady-state noise for EAS hearing, and the access to the low-frequency acoustic regions provides an avenue for investigating other techniques

such as ideal binary masking that directly target glimpsing improvement (Wang, 2005; Brungart *et al.*, 2006; Hu and Loizou, 2008). In addition, it is worthwhile to investigate how to integrate noise reduction into EAS processing more efficiently, as noise reduction is carried out in both hearing aids and cochlear implants at the same time.

The present study demonstrated that regenerating harmonics with the associated sidebands could provide significant benefits for EAS hearing in noise. In some real-world applications, the low-frequency portions where F0 resides could be missing. For instance, telephone speech is characterized with a passband between 300 Hz and 3400 Hz due to the transmission network (Hu and Loizou, 2007b, Fig. 2); under these circumstances, regenerating the missing low-numbered harmonics and the associated sidebands may enhance the F0 representation. Regenerating harmonics has been proven to be beneficial not only for EAS processing (as demonstrated in the present study) but also for vocoder processing in the absence of low-frequency acoustic information (Hu and Loizou, 2010).

The deterministic sideband constraint implemented in the present study is very crude, and it would be interesting to examine if a substituted temporal envelope extracted from the speech signal can provide additional benefits. Also, the sideband constraint can be used as a low computational-cost cue to improve EAS hearing in noise as sideband shape generally degrades in noise. For instance, at a higher SNR level, moderate noise and more accurate noise estimation probably keep the harmonic structures relatively intact compared with the clean speech, but they degrade the sideband shape. The results in Fig. 5 show that the sideband constraint is beneficial. This warrants investigating whether applying the sideband constraint without harmonics regeneration after noise reduction (or even directly to the noisy signal) is beneficial; if so, it can serve as a low-cost, post-processing step.

The present study used speech stimuli produced by a male speaker, and the impact of gender remains unclear in the evaluated strategies. Female talkers typically have higher F0s than male talkers, and the available low-numbered harmonics are fewer in the low-frequency acoustic regions. If the F0 is greater than 300, in the present experiment setup, there is only one usable harmonic at the F0. Under these circumstance, there are several alternative strategies. First, the benefits to extend the harmonics regeneration from 600 Hz up to more than 3200 Hz can be investigated, which probably can improve representation of other formants (such as F2/F3) in the high-frequency temporal envelopes. Second, harmonics have wider spacing in female speech, and the benefits of better sideband constraint modeling can be investigated.

Finally, Brown and Bacon (2009) proposed the use of a tone carrying F0 and envelope cues as a beneficial strategy. The present experiment results suggest that a viable alternative strategy is to use the harmonics to generate the temporal fine structures that are modulated by a deterministic temporal envelope or a temporal envelope extracted from speech signal.

Several of the strategies described above are currently under investigation in our laboratory.

ACKNOWLEDGMENTS

This research was supported by NIH/NIDCD under Grant No. R03-DC008887. The author gratefully acknowledges Dr. Philipos C. Loizou for many helpful discussions about this work. The author also thanks Dr. Ruth Litovsky and two reviewers for their very helpful comments, and Michelle M. Schoenecker for her editorial assistance.

- Brown, C. A., and Bacon, S. P. (2009). "Low-frequency speech cues and simulated electric-acoustic hearing," *J. Acoust. Soc. Am.* **125**, 1658–1665.
- Brungart, D., Chang, P., Simpson, B., and Wang, D. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," *J. Acoust. Soc. Am.* **120**, 4007–4018.
- Chang, J. E., Bai, J. Y., and Zeng, F. (2006). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," *IEEE Trans. Biomed. Eng.* **53**, 2598–2601.
- Chu, W. C. (2004). "Vector quantization of harmonic magnitudes in speech coding applications—A survey and new technique," *EURASIP J. Appl. Signal Process.* **2004**, 2601–2613.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.
- Dorman, M. F., Gifford, R. H., Spahr, A. J., and McKarns, S. A. (2008). "The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies," *Audiol. Neuro-Otol.* **13**, 105–112.
- 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," *Ear Hear.* **26**, 371–380.
- Ephraim, Y., and Malah, D. (1985). "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.* **33**, 443–445.
- Gantz, B. J., Turner, C. W., and Gfeller, K. E. (2006). "Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/nucleus hybrid implant," *Audiol. Neuro-Otol.* **11**, 63–68.
- Gantz, B. J., Turner, C. W., Gfeller, K. E., and Lowder, M. W. (2005). "Preservation of hearing in cochlear implant surgery: Advantages of combined electrical and acoustical speech processing," *Laryngoscope* **115**, 796–802.
- Glasberg, B., and Moore, B. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Hu, Y., and Loizou, P. (2007a). "A comparative intelligibility study of single-microphone noise reduction algorithms," *J. Acoust. Soc. Am.* **122**, 1777–1786.
- Hu, Y., and Loizou, P. (2007b). "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Commun.* **49**, 588–601.
- Hu, Y., and Loizou, P. (2008). "A new sound coding strategy for suppressing noise in cochlear implants," *J. Acoust. Soc. Am.* **124**, 498–509.
- Hu, Y., and Loizou, P. (2009). "Toward restoring speech intelligibility in noise," in *Proceedings of the 2009 Conference on Implantable Auditory Prostheses*, Oral Presentation, Lake Tahoe, California.
- Hu, Y., and Loizou, P. (2010). "On the importance of preserving the harmonics and neighboring partials prior to vocoder processing: Implications for cochlear implants," *J. Acoust. Soc. Am.* **127**, 427–434.
- Hu, Y., Loizou, P., Li, N., and Kasturi, K. (2007). "Use of a sigmoidal-shaped function for noise attenuation in cochlear implant," *J. Acoust. Soc. Am.* **122**, EL128–EL134.
- Huang, X., Acero, A., and Hon, H.-W. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm and System Development* (Prentice-Hall, Englewood Cliffs, NJ).
- IEEE (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Kiefer, J., Pok, M., Adunka, O., Stürzebecher, E., Baumgartner, W., Schmidt, M., Tillein, J., Ye, Q., and Gstöettner, W. (2005). "Acoustic simulations of combined electric and acoustic hearing," *Audiol. Neuro-Otol.* **10**, 134–144.
- Kim, G., Lu, Y., Hu, Y., and Loizou, P. (2009). "An algorithm that improves speech intelligibility in noise," *J. Acoust. Soc. Am.* **126**, 1486–1494.
- Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.
- Kondoz, A. M. (1999). *Digital Speech: Coding for Low Bit Rate Communication Systems* (Wiley, New York).
- Kong, Y., and Carlyon, R. P. (2007). "Improved speech recognition in noise

- in simulated binaurally combined acoustic and electric stimulation,” *J. Acoust. Soc. Am.* **121**, 3717–3727.
- Kong, Y., Stickney, G. S., and Zeng, F. (2005). “Speech and melody recognition in binaurally combined acoustic and electric stimulation,” *J. Acoust. Soc. Am.* **117**, 1351–1361.
- Li, N., and Loizou, P. C. (2007). “Factors influencing glimpsing of speech in noise,” *J. Acoust. Soc. Am.* **122**, 1165–1172.
- Li, N., and Loizou, P. C. (2008). “A glimpsing account for the benefits of simulated combined acoustic and electric hearing,” *J. Acoust. Soc. Am.* **123**, 2287–2294.
- Loizou, P. (2007). *Speech Enhancement: Theory and Practice* (Taylor & Francis, FL).
- Loizou, P., Lobo, A., and Hu, Y. (2005). “Subspace algorithms for noise reduction in cochlear implants,” *J. Acoust. Soc. Am.* **118**, 2791–2793.
- Mines, M., Hanson, B., and Shoup, J. (1978). “Frequency of occurrence of phonemes in conversational English,” *Lang Speech* **21**, 221–241.
- Mitra, S. K. (2001). *Digital Signal Processing—A Computer-Based Approach* (McGraw-Hill, New York).
- Moore, B. C. J., and Carlyon, R. P. (2005). “Perception of pitch by people with cochlear hearing loss and by cochlear implant users,” in *Pitch: Neural Coding and Perception* (Springer, Berlin), pp. 234–277.
- Nehorai, A., and Porat, B. (1986). “Adaptive comb filtering for harmonic signal enhancement,” *IEEE Trans. Acoust., Speech, Signal Process.* **34**, 1124–1138.
- Plack, C. J., and Oxenham, A. J. (2005). “The psychophysics of pitch,” in *Pitch: Neural Coding and Perception* (Springer, Berlin), pp. 7–55.
- Plapous, C., Marro, C., and Scalart, P. (2006). “Improved signal-to-noise ratio estimation for speech enhancement,” *IEEE Trans. Audio, Speech, Lang. Process.* **14**, 2098–2108.
- 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.
- Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). “Speech recognition with primarily temporal cues,” *Science* **270**, 303–304.
- Stylianou, Y. (2001). “Applying the harmonic plus noise model in concatenative speech synthesis,” *IEEE Trans. Speech Audio Process.* **9**, 21–29.
- 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.
- vonIlberg, C., Kiefer, C., Tillein, J., Pfenningdorff, T., Hartman, R., Sturzebecher, E., and Klinke, R. (1999). “Electric-acoustic stimulation of the auditory system,” *ORL* **61**, 334–340.
- Wang, D. (2005). “On ideal binary mask as the computational goal of auditory scene analysis,” *Speech Separation by Humans and Machines* (Kluwer, Dordrecht), pp. 181–197.
- Yang, L., and Fu, Q. (2005). “Spectral subtraction-based speech enhancement for cochlear implant patients in background noise,” *J. Acoust. Soc. Am.* **117**, 1001–1003.
- Zakis, J. A., McDermott, H. J., and Vandali, A. E. (2007). “A fundamental frequency estimator for the real-time processing of musical sounds for cochlear implants,” *Speech Commun.* **49**, 113–122.
- Zavarehei, E., Vaseghi, S., and Yan, Q. (2007). “Noisy speech enhancement using harmonic-noise model and codebook-based post-processing,” *IEEE Trans. Audio, Speech, Lang. Process.* **15**, 1194–1203.