Speech Perception With Combined Electric-Acoustic Stimulation: A Simulation and Model Comparison

Tobias Rader,[1,2] Youssef Adel,[1,2] Hugo Fastl,[2] and Uwe Baumann[1]

Objective: The aim of this study is to simulate speech perception with combined electric-acoustic stimulation (EAS), verify the advantage of combined stimulation in normal-hearing (NH) subjects, and then compare it with cochlear implant (CI) and EAS user results from the authors' previous study. Furthermore, an automatic speech recognition (ASR) system was built to examine the impact of low-frequency information and is proposed as an applied model to study different hypotheses of the combined-stimulation advantage. Signal-detection-theory (SDT) models were applied to assess predictions of subject performance without the need to assume any synergistic effects.

Design: Speech perception was tested using a closed-set matrix test (Oldenburg sentence test), and its speech material was processed to simulate CI and EAS hearing. A total of 43 NH subjects and a customized ASR system were tested. CI hearing was simulated by an aurally adequate signal spectrum analysis and representation, the part-tone-time-pattern, which was vocoded at 12 center frequencies according to the MED-EL DUET speech processor. Residual acoustic hearing was simulated by low-pass (LP)-filtered speech with cutoff frequencies 200 and 500 Hz for NH subjects and in the range from 100 to 500 Hz for the ASR system. Speech reception thresholds were determined in amplitude-modulated noise and in pseudocontinuous noise. Previously proposed SDT models were lastly applied to predict NH subject performance with EAS simulations.

Results: NH subjects tested with EAS simulations demonstrated the combined-stimulation advantage. Increasing the LP cutoff frequency from 200 to 500 Hz significantly improved speech reception thresholds in both noise conditions. In continuous noise, CI and EAS users showed generally better performance than NH subjects tested with simulations. In modulated noise, performance was comparable except for the EAS at cutoff frequency 500 Hz where NH subject performance was superior. The ASR system showed similar behavior to NH subjects despite a positive signal-to-noise ratio shift for both noise conditions, while demonstrating the synergistic effect for cutoff frequencies ≥300 Hz. One SDT model largely predicted the combined-stimulation results in continuous noise, while falling short of predicting performance observed in modulated noise.

Conclusions: The presented simulation was able to demonstrate the combined-stimulation advantage for NH subjects as observed in EAS users. Only NH subjects tested with EAS simulations were able to take advantage of the gap listening effect, while CI and EAS user performance was consistently degraded in modulated noise compared with performance in continuous noise. The application of ASR systems seems feasible to assess the impact of different signal processing strategies on speech perception with CI and EAS simulations. In continuous noise, SDT models were largely able to predict the performance gain without assuming any synergistic effects, but model amendments are required to explain the gap listening effect in modulated noise.

Key words: Automatic speech recognition, Cochlear implant, Electric-acoustic stimulation, Modeling, Signal detection theory, Simulation, Speech perception.

(Ear & Hearing 2015;XX;00–00)

INTRODUCTION

Combined electric-acoustic stimulation (EAS) comprises electrical stimulation by a cochlear implant (CI) and acoustic stimulation of low-frequency residual hearing in the same ear. First described by von Ilberg et al. (1999), EAS was made possible by advances in the surgical approach and electrode arrays designed to minimize trauma of the delicate cochlear structures, thus allowing the preservation of acoustic low-frequency hearing after implantation (Gantz & Turner 2003; Gantz & Turner 2004; Helbig et al. 2011; review in von Ilberg et al. 2011). EAS is now established as a therapeutic treatment for people with severe-to-profound hearing loss but remaining low-frequency hearing although optimizations of its clinical application are still being discussed (Gifford & Dorman 2012; Incerti et al. 2013).

EAS users generally have significantly better scores on speech intelligibility tests, especially in complex noise situations, than do bilateral CI users (review in Dorman & Gifford 2010; Rader et al. 2013). This effect is often described as the "combined-stimulation advantage" and was generally studied by direct comparison of CI and EAS user performance or by using simulations with normal-hearing (NH) subjects. Furthermore, the underlying mechanisms were discussed on the basis of perceptual models, primarily assessing the apparently synergistic effect of EAS and how it could be explained.

Different modes of combining electric and acoustic stimulation must be distinguished when comparing user performance. Apart from combined electric and acoustic stimulation in the same ear (henceforth called EAS), some people receive a standard CI with a fully inserted electrode and a hearing aid on the contralateral ear, which is termed bimodal stimulation. The latter group was also shown to benefit from low-frequency acoustic hearing in the contralateral ear (e.g., Gifford et al. 2007; Cullington & Zeng 2010). While EAS and bimodal users receive different types of information (Ching et al. 2007), both groups outperform CI users in speech intelligibility tests (Dorman & Gifford 2010). In our previous study (Rader et al. 2013), performance of EAS users with a hearing aid in the contralateral ear (bimodal EAS) was compared with performance of bilateral CI users using the Oldenburg sentence test (OLSA; Wagener et al. 1999a, 1999b, 1999c) in a multisource noise field. Gap listening (or "glimpsing"), bilateral interactions, and frequency fine structure were discussed with regard to the benefit of EAS. The results indicated that binaural interaction between EAS and contralateral acoustic hearing enhances speech perception in complex noise situations similar to a cocktail party scenario. However, neither bilateral CI nor bimodal EAS users were able to benefit from gap

[1]Department of Audiological Acoustics, ENT Department, University Hospital Frankfurt, Frankfurt, Germany; and [2]Arbeitsgruppe Technische Akustik, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, Munich, Germany.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

0196/0202/15/XXXX-0000/0 • Ear & Hearing • Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

<zdoi; 10.1097/AUD.0000000000000178>

listening, contrary to what was observed in NH subjects listening to acoustic simulations (Li & Loizou 2008).

To investigate the psychophysical mechanisms of the combined-stimulation advantage, NH subjects were tested with EAS simulations, which have consistently demonstrated the combined-stimulation advantage in comparison with CI simulation (i.e., vocoded speech). This advantage was mainly investigated by evaluating speech intelligibility in noise and inspecting the role of fundamental frequency (F0) and temporal fine-structure information (Qin & Oxenham 2006; Kong & Carlyon 2007; Brown & Bacon 2010). It was suggested that low-frequency speech signals provide cues to differentiate target-related and masker-related information, which results in a synergistic effect of combining low-frequency and vocoded speech signals. In a study comparing different models of the combined-stimulation advantage, Micheyl and Oxenham (2012) referred to this as the "super-additivity hypothesis." A different supposition was that important phonetic cues in low-frequency speech signals can supplement limited cues in vocoded speech (Qin & Oxenham 2006; Kong & Carlyon 2007; Brown & Bacon 2010). Micheyl and Oxenham (2012) therefore suggested that having access to two sources of information can account for the combined-stimulation advantage, without the need for synergistic interactions between low-frequency and vocoded speech. They referred to this as the "simple-additivity hypothesis."

Mathematical and perceptual models have therefore been applied to validate different psychophysical hypotheses. To predict subject performance, Kong and Carlyon (2007) used the probability-summation model, which was derived from the probability of occurrence of either or both of two independent and not mutually exclusive events, i.e., identifying low-frequency and vocoded speech. This model, however, typically underpredicted subject performance. On the other hand, signal-detection-theory (SDT) models were largely able to predict subject performance, without assuming any synergistic effects from combining low-frequency and vocoded speech (Seldran et al. 2011; Micheyl & Oxenham 2012).

In contrast to perceptual models that are commonly studied to explain the psychophysical processes, automatic speech recognition (ASR) is aimed to provide a human–machine interface that translates spoken words into text. ASR systems convert relevant spectral information into quantifiable acoustic features to reduce speech signal variability and ensure robust ASR performance. Such acoustic features are often based on the spectral envelope of the short-term speech spectrum, e.g., mel-frequency cepstral coefficients. Cong-Thanh Do et al. (2010) used spectrally reduced speech, which was synthesized from a filter bank simulating the motion of the basilar membrane, to evaluate hidden Markov model (HMM)-based ASR. Results showed that 16 frequency subbands were sufficient for the ASR system to achieve significantly high word recognition accuracy. This demonstrated that spectrally reduced speech can yield sufficient information for stable speech recognition, thus motivating the application of ASR to further inspect the synergistic effect of EAS.

In conclusion, the combined-stimulation advantage is well established and supported by the majority of data comparing CI with EAS user performance. However, its psychophysical mechanism remains to be explained and its principal components need to be determined. In the present study, we established a simulation of speech perception with EAS to verify the synergistic effect in NH subjects and compared it with results of bilateral CI and bimodal EAS users from our previous study (Rader et al. 2013). To this end, the simulation was based on OLSA speech material and tested in pseudocontinuous and in amplitude-modulated noise. We then used two approaches, an applied ASR model and two SDT models, to further explain the performance gain observed in our EAS simulations. The ASR system was built to investigate the impact of low-frequency information conveyed by F0 and lower harmonics on the recognition of spectrally reduced speech; it is hereby proposed as an applied model to isolate psychophysical parameters and study different hypotheses of the combined-stimulation advantage. SDT model predictions were compared with our simulation data to discuss the apparently synergistic effect of combined stimulation.

MATERIALS AND METHODS

NH Subjects

Data were collected in three separate sessions with different sets of subjects, with a total n = 43 (ages between 22 and 27 years). Speech tests were conducted with unprocessed speech in a control group (n = 21) and with simulated speech in a different group (n = 22). All subjects were native German language speakers and had NH, which was defined as having pure-tone audiogram deviances ≤20 dB relative to the standard hearing threshold (Zwicker & Heinz 1955; ISO 7029:2001) over the frequency range 0.2 to 24 Bark (20–15,500 Hz). Audiograms were measured using Bekesy tracking as implemented by Seeber et al. (2003). All sessions were conducted at the Institute for Human-Machine Communication of the Technical University in Munich as part of a workshop for students, thus sample sizes for different experimental conditions depended on student attendance.

ASR System

The ASR model approach was built using the HMM Toolkit (HTK, version 3.4.1; University of Cambridge), which was mainly designed for HMM-based speech processing tools (Young et al. 2006). For this purpose, HTK assumes that speech is a sequence of symbols in which language is encoded. A stream of continuous speech is regarded as stationary for intervals of about 10 msec and thus divided into short time segments. These segments are each coded as parametrical speech vectors that constitute an observation sequence in series. This abstraction allows an ASR system to be implemented based on discrete statistical models such as HMMs. Multiple utterances of the same speech data (e.g., phonemes) are assigned to a Markov chain and then the probability of each observation sequence being generated by that chain is algorithmically maximized and hence the adaption of the model parameters to the speech vectors. For an unknown observation sequence, the likelihood of each chain generating that sequence is computed and the most likely is translated back to the corresponding speech symbols. This study incorporates recommended parameters to construct a small vocabulary continuous speech recognizer in HTK (Vertanen 2006; Young et al. 2006), while omitting refinements not applicable to the German language.

The ASR system was trained and tested using speech data and a given task. Here, the task grammar of the closed-set Oldenburg sentence test (OLSA) was defined and used for training context-dependent three-state left-to-right triphone HMMs. The unprocessed speech data of OLSA consisted of 20 lists with 30 sentences each, which were used in conjunction with a pronunciation dictionary to provide phone-level transcriptions needed to initialize the HMM training process. For this purpose, a comprehensive German language pronunciation dictionary was used, which was part of a work concerning large vocabulary continuous speech recognition based on the machine-readable phonetic alphabet SAMPA (Wells 1997; Weninger et al. 2010). Speech data were coded as mel-frequency cepstral coefficient vectors using a mel-scale filter bank with 26 channels limited to 8 kHz. This corresponds to the upper cutoff frequency of the DUET speech processor, with which EAS users in the previous study were fitted (Rader et al. 2013). The robustness of trained HMMs relies, for the most part, on the phoneme distribution of speech data. If a phoneme does not occur frequently, then the corresponding HMM is more susceptible to mistakes. Although OLSA speech data were unevenly distributed, the closed-set sentence test provided a consistent basis for comparison. In addition, ASR evaluation using unprocessed OLSA in quiet yielded 99.9% words correct, therefore validating the model for the specific study purpose.

Speech Material

OLSA was used to test speech intelligibility (see Speech Test), and its speech material was used to simulate CI and EAS speech perception. The general schematic of the applied signal processing is shown in Figure 1 and is elaborated below.

CI Simulation • OLSA speech material was processed using an aurally adequate signal spectrum analysis and representation, the part-tone-time-pattern (PTTP; Heinbach 1988; Baumann 1995). First, a Fourier-t-transformation (Terhardt 1985) was carried out with analysis bandwidths proportional to the critical bands of the ear. Second, frequency and phase were analyzed and a magnitude spectrum was calculated. Last, maxima detection was conducted to obtain the PTTP of the input signal. The graphical PTTP signal representation (or “maxigram”) of the German sentence “Stefan gewann zwölf grüne Blumen” (Stefan won twelve green flowers) is shown in Figure 2A. To simulate hearing by means of a CI, the continuous frequency spectrum was divided into 12 band-pass filter channels with cutoff frequencies corresponding to these of the DUET speech processor (Table 1). The filter bank was limited to 8 kHz with a lower cutoff frequency of 500 Hz, which is the typical frequency range that most people with a DUET speech processor use. Resynthesis was then realized using a 12-band sinusoidal vocoder that matched signal phase at sampling points, that is, with phasing continuity. Vocoder frequencies were set to the respective channel center frequencies of the DUET speech processor filter bank, which were determined by the fitting

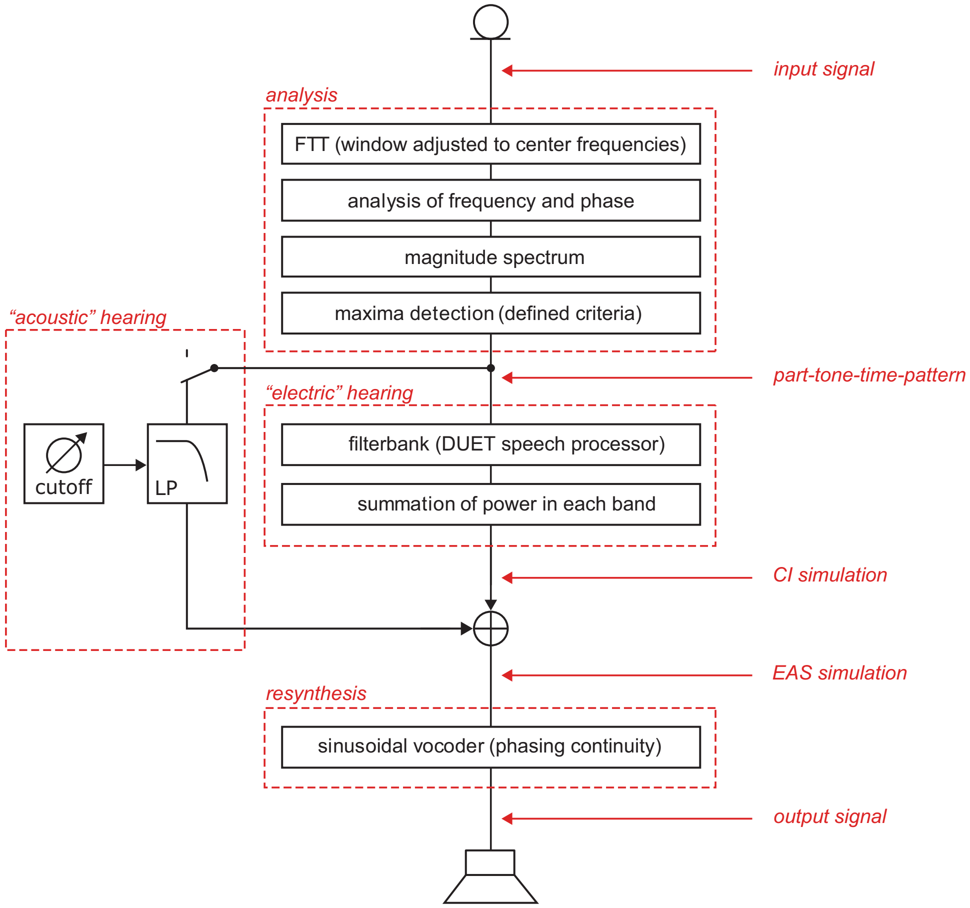
input signal
analysis
FTT (window adjusted to center frequencies)
analysis of frequency and phase
magnitude spectrum
maxima detection (defined criteria)
“acoustic” hearing
part-tone-time-pattern
“electric” hearing
cutoff
LP
filterbank (DUET speech processor)
summation of power in each band
CI simulation
EAS simulation
resynthesis
sinusoidal vocoder (phasing continuity)
output signal

Fig. 1. Schematic of signal processing to simulate cochlear implant (CI) and electric-acoustic stimulation (EAS) hearing using the part-tone-time-pattern (PTTP). First, the input signal was analyzed and converted to the PTTP domain. For CI simulation, pitch information of each frequency band of the DUET speech processor (Table 1) was collapsed and then mapped to the respective center frequency. For EAS with residual acoustic low-frequency hearing, the low-pass (LP)-filtered PTTP signal with variable cutoff frequency was added to the CI simulation. Finally, the output signal was generated using resynthesis with a sinusoidal vocoder. FTT indicates Fourier-t-transformation.

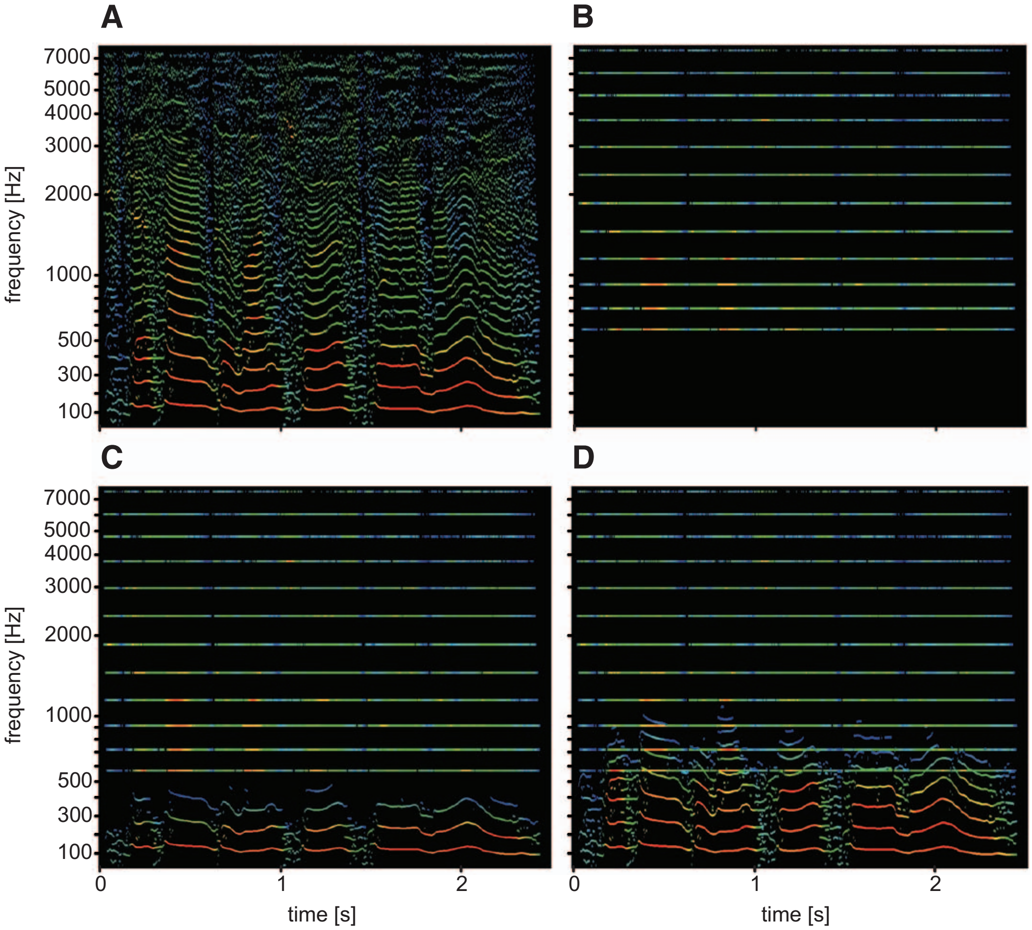

Fig. 2. Maxigram (graphic representation of the part-tone-time-pattern, part-tone level is color-coded from low in blue to high in red) of the German sentence "Stefan gewann zwölf grüne Blumen" (Stefan won twelve green flowers). A, Unprocessed signal. B, Cochlear implant simulation. C, Electric-acoustic simulation with low-pass (LP) cutoff frequency f_{LP} = 200 Hz. D, electric-acoustic simulation with f_{LP} = 500 Hz.

software CI.STUDIO+ (MED-EL, Innsbruck, Austria). It is safe to assume that no sidebands were formed in the resynthesis although amplitude fluctuations might occur if several harmonics fall into the same frequency band of a respective part-tone component. Nevertheless, part tones are not synchronously resynthesized so that the final signal of each part tone would show no modulation in the F0 range. The resulting signal simulated CI-processed speech (i.e., electric hearing) containing only amplitude information at the 12 channels, but without fine-structure frequency information. This is comparable with the continuous interleaved sampling strategy of CIs. The corresponding PTTP signal representation is shown in Figure 2B.

EAS Simulation • To simulate EAS speech perception, additional speech information must be provided as per low-frequency residual hearing (i.e., acoustic hearing). To this end, the speech signal was low-pass (LP) filtered in the PTTP domain at cutoff frequencies f_{LP} = 100 to 500 Hz in steps of 100 Hz with a slope of −60 dB per octave to map different degrees of residual hearing (e.g., LP_{100} for f_{LP} = 100 Hz). Adding this signal to the CI simulation allowed for EAS with different amounts of low-frequency information to be simulated, where the cutoff frequency of the LP-filtered speech is indicated by a subscript (e.g., EAS_{100} for the addition of LP_{100} to the CI simulation). Figure 2C and D shows the PTTP signal representation for EAS_{200} and EAS_{500}, respectively.

Audio samples, which are provided in Supplemental Digital Content 1 (http://links.lww.com/EANDH/A190), demonstrate the LP_{500}, CI simulation, EAS_{500} simulation, and the unprocessed signal conditions for the English sentence "Thomas wants nine cheap bags" (original sample from Zokoll et al. 2013).

Noise Characteristics

Two types of competing noise signals were used in the speech tests:

1. Pseudocontinuous noise of the Oldenburg sentence test (OL-noise): this noise signal is generated by an averaging process of 30 time-shifted OLSA test sentences (Wagener et al. 1999a, 1999b, 1999c). The frequency power spectrum closely resembles the short-term power spectrum of each OLSA test sentence and is restricted to frequencies below 12.6 kHz. In addition, temporal modulation is nearly absent, and cues for gap listening are not available. As a result of the high masking efficacy of OL-noise, the speech discrimination function is very steeply sloped (17.1% per 1 dB) and allows efficient and exact estimations of individual speech reception thresholds (SRT).
2. Amplitude-modulated Fastl-noise: based on the noise signal developed by the "Comité Consultatif International

Télégraphique et Téléphonique" noise (according to ITU-T Rec. G.227 [11/88] conventional telephone signal), Fastl-noise aims to represent the temporal characteristics of speech (Fastl 1987). Comité Consultatif International Télégraphique et Téléphonique noise was amplitude modulated at a randomly varying modulation frequency with a distribution peak at 4 Hz, correlating with the amplitude modulation statistics of the German language. Fastl-noise offers no informational masking and, because of its slow temporal modulation, provides the opportunity to take advantage of gap listening (or "glimpsing"). Consequently, the speech discrimination function has a flatter slope with Fastl-noise—an estimated 8% per 1 dB—than with OL-noise (Fastl & Zwicker 2007).

Experimental Setup

Tests with NH subjects were conducted using a computer equipped with a high-quality 24-bit 8-channel AD/DA converter and a headphone amplifier (RME Hammerfall DSP Multiface II). Stimuli were presented binaurally using circumaural headphones (Sennheiser HDA 200). The test was run by a MATLAB (The MathWorks, Inc.) procedure with a graphical user interface for subject feedback. The speech signal was presented at a fixed sound pressure level (SPL) of 65 dB; the noise level was adaptively adjusted depending on the number of correctly discriminated words in the sentence. Calibration was accomplished in reference to dB SPL with a B&K (Brüel & Kjær, Nærum, Denmark) 4153 artificial ear and a B&K 0.5-inch 4134 microphone, a B&K 2669 preamplifier, a B&K 2690 measuring amplifier, and an NTi (NTi Audio AG, Schaan, Liechtenstein) AL1 sound level meter.

An equivalent procedure was implemented for the ASR system using a MATLAB script conjoined with the software interface of HTK. The same stimuli used with NH subjects were given as input to the ASR system, and the recognized words were fed back to the adaptive test procedure.

Speech Test

The closed-set OLSA was used to test speech intelligibility of NH subjects and the ASR system. In the closed-set mode, subjects indicated which speech items were understood on a touchscreen (5 words and 10 possibilities per word). Consequently, the options were available to subjects in advance. The adaptive test procedure yielded SRTs defined as the signal-to-noise ratio (SNR) at 50% correct speech recognition. The speech signal was presented at a fixed level of 65 dB SPL; the noise level was adaptively adjusted depending on the number of correctly recognized words in the sentence from each trial. Initial step sizes were 5, 3, 1, −1, −3, and −5 dB SNR for 0, 1, 2, 3, 4, and 5 correct words, respectively. The starting SNR was set to 0 or 5 dB, and the step sizes were adjusted according to the number of reversals as suggested by Brand and Kollmeier (2002). The SRT was then calculated as the mean of the last 10 reversals. Further details are provided in our previous study (Rader et al. 2013).

TABLE 1. Channel numbers and cutoff frequencies of the filters used by default in maps with lower cutoff frequency set to 500 Hz in the MED-EL DUET speech processor and corresponding center frequencies used for resynthesis of the CI simulation

Channel Number	Cutoff Frequency (Hz) Lower	Upper	Resynthesis Frequency (Hz)
1	500	637	567
2	638	807	717
3	808	1022	909
4	1023	1294	1150
5	1295	1639	1457
6	1640	2076	1845
7	2077	2628	2336
8	2629	3328	2958
9	3329	4215	3746
10	4216	5337	4743
11	5338	6759	6007
12	6760	7999	7606

CI, cochlear implant; EAS, electric-acoustic stimulation.

NH Subjects • Speech tests for NH subjects were conducted in the following conditions: (1) unprocessed, (2) CI simulation, (3) LP-filtered speech LP_{200} and LP_{500}, and (4) EAS_{200} and EAS_{500} simulations (see Speech Material). The unprocessed condition indicates an OLSA test with its original speech material, which served as a control. The CI simulation then assessed subject performance with spectrally reduced speech. The LP-filtered speech was tested as a reference for the information provided by lower harmonics alone. Finally, the impact of adding LP-filtered speech to the CI simulation was examined using EAS simulations at different cutoff frequencies. NH subjects were trained in each condition with a single OLSA list of 20 sentences in the respective condition before each test.

ASR System • The ASR system was also tested in the aforementioned conditions, that is, (1) unprocessed, (2) CI simulation, (3) LP-filtered speech LP_{100} to LP_{500}, and (4) EAS_{100} to EAS_{500} simulations, whereas the cutoff frequencies were set to f_{LP} = 100 to 500 Hz in steps of 100 Hz (see Speech Material). Because the test was automated, it was possible to examine a larger set of parameters, while still limited by the extended processing time due to the adaptive procedure. In contrast to NH subjects, the ASR system can be viewed as a deterministic system for a given set of speech material; thus, evaluations of the ASR system were also conducted at fixed SNR levels. This was done by computing the total percentage of correct word recognition for 20 OLSA lists, with 30 sentences each.

RESULTS

Speech Discrimination Functions

As a control, 21 NH subjects were tested with unprocessed OLSA speech material. In continuous OL-noise, the mean SRT was −6.6 dB SNR with standard deviation ±0.8 dB SNR comparable with the OLSA evaluation data (Wagener et al. 1999a). In modulated Fastl-noise, the mean SRT was −14.3 dB SNR with standard deviation ±3.5 dB SNR.

Speech discrimination functions of the ASR system were computed using unprocessed OLSA speech material and are plotted in Figure 3. Performance improved monotonically with increasing SNR in both noise conditions. In continuous noise, performance increased from 9.4% to 99% correct in the range from −7 to +12 dB SNR, with an SRT level of 3.4 dB SNR. In modulated noise, the discrimination function exhibited a negative SNR shift of 10.9 dB, with an SRT level of −7.5 dB SNR. As with NH subjects, modulated noise evidently allowed gap listening, enabling the ASR system to achieve a better SRT than in continuous noise. This 10.9-dB SNR improvement is comparable with the mean 7.7-dB SNR improvement for NH subjects.

Linear slope approximations yielded 8.4% per dB SNR in continuous noise and 6.6% per dB SNR in modulated noise. Compared with slopes measured for NH subjects (17.1% per dB SNR in modulated and 8% per dB SNR in continuous noise), the speech discrimination functions are flatter for both noise conditions. In addition, the relatively steep slope measured for NH subjects in continuous OL-noise (Wagener et al. 1999a) was not observed, which results in degraded SRT measurement precision for the ASR system.

CI and EAS Simulation

For NH subjects, SRT results of CI, EAS_{200}, and EAS_{500} simulations are shown in Figure 4 and compared with bilateral CI and bimodal EAS user results from the study by Rader et al. (2013).

For the CI simulation in modulated noise, median SRT was 6.2 dB SNR compared with 6.8 dB SNR for CI users, with no significantly different distributions ($p = 0.69$, Student's *t* test). In continuous noise, however, median SRT was 5.1 dB SNR for CI simulation and −0.4 dB SNR for CI users, with significantly different distributions ($p < 0.001$). While CI user performance improved in continuous noise, performance of NH subjects listening to CI simulation showed no considerable difference between each noise condition.

Testing with LP-filtered speech did not yield valid SRT scores for LP_{200} (n = 8, range, 21–41.5 dB SNR). Limited results for LP_{500} (n = 4) had median SRT at 2.4 dB SNR in modulated noise and 3.15 dB SNR in continuous noise, which are not shown here.

The addition of LP-filtered speech resulted in significant SRT improvement when compared with CI simulation results; in modulated noise, median SRT was −0.4 dB SNR for EAS_{200} and −10.1 dB SNR for EAS_{500}. In continuous noise, median SRT was 2.1 dB SNR for EAS_{200} and −1.0 dB SNR for EAS_{500}. Similarly, EAS users performed significantly better than CI users in modulated ($p < 0.001$) and in continuous ($p = 0.031$) noise. For NH subjects in modulated noise, results of EAS_{200} simulation were in fair agreement with EAS users, yet results of EAS_{500} simulation were significantly better ($p < 0.001$). In continuous noise, results of EAS_{200} simulation were significantly worse ($p < 0.001$) than EAS users, while EAS_{500} simulation results were comparable.

ASR System Performance

Speech recognition scores of the ASR system in quiet and at a fixed 0-dB SNR level for both noise conditions are shown in Figure 5. Scores are given as percentages of correctly recognized words for 20 OLSA lists, with 30 sentences each.

For the CI simulation, a score of 26.3% was achieved in quiet, 16.3% in modulated noise, and 13.5% in continuous noise. As the speech recognition score in quiet was well below 50%, a corresponding SRT level could not be determined. Likewise, testing the ASR system with LP-filtered speech in quiet yielded a score of 11.4% for LP_{200} and 22.3% for LP_{500}. Consequently, SRT levels could not be determined for either of these conditions. To underline the poor performance in these conditions, consider that the task grammar of the closed-set OLSA was predefined (see ASR System); therefore, the probability of correctly recognizing a word by chance was 10% (i.e., 10 possibilities for each word).

While a score of only 32.5% was achieved for the EAS_{100} simulation in quiet, clear improvements were observed for the EAS_{200} to EAS_{500} simulations and reached 99.5% for EAS_{300} in quiet. Compared with 99.9% for the unprocessed condition, the positive effect of adding LP-filtered speech on speech recognition stability was evident. This was also observed in modulated noise at 0 dB SNR although the score in continuous noise was limited to 23.4% at 0 dB SNR for the unprocessed condition. Consistent with the results of EAS users and NH subjects tested with simulations, the "super-additive" effect was observed in the ASR results, where the recognition score for the combined signal exceeded the sum of scores for the CI simulation and the LP condition.

For a direct comparison with NH simulation SRTs (Fig. 4), thresholds from an adaptive procedure in accordance with OLSA and scores at a large range of fixed SNR levels were computed. Corresponding ASR system SRTs can be found in

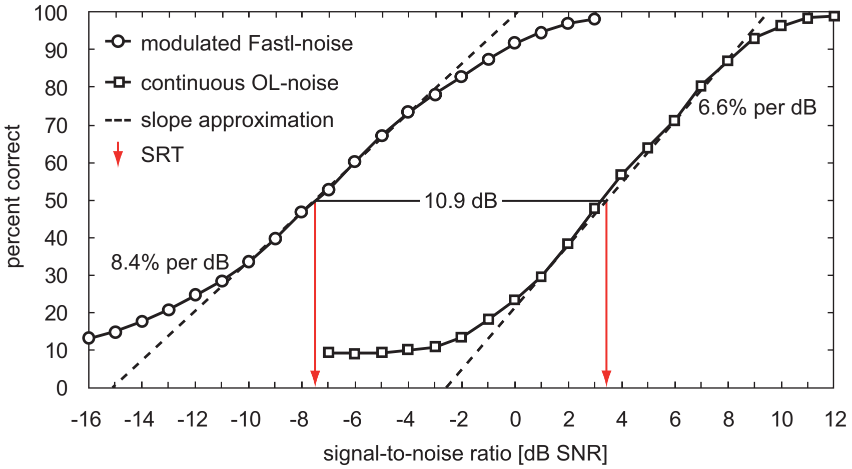

Fig. 3. Speech discrimination functions of the automatic speech recognition system for modulated Fastl-noise (circles) and continuous OL-noise (squares). Arrows indicate the speech reception threshold (SRT) levels for each function, and dashed lines are corresponding linear slope approximations. SNR indicates signal-to-noise ratio.

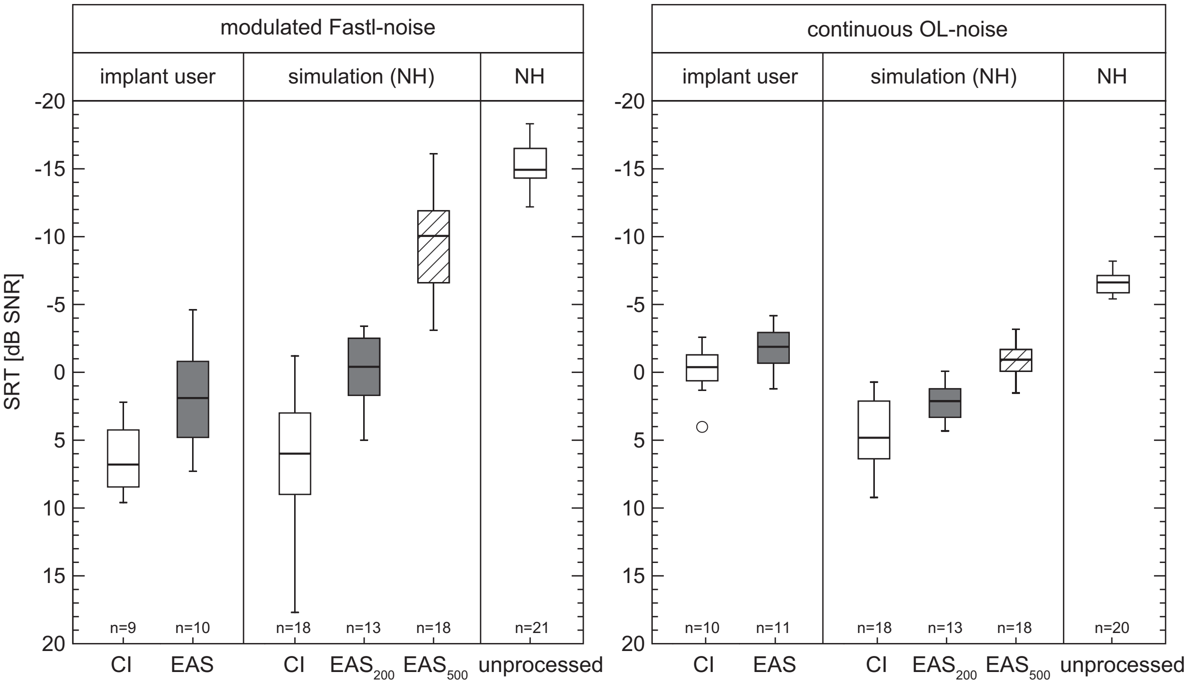

Fig. 4. Speech reception threshold (SRT) box plots (i.e., median, 1st and 3rd quartiles, minimum and maximum values, circles indicate outliers, n denotes number of subjects for each condition) in modulated Fastl-noise (left) and continuous OL-noise (right). The SRT axis is reversed, thus better performance corresponds to higher positioning on the graph. Results of bilateral cochlear implant (CI) and bimodal electric-acoustic stimulation (EAS) users from the study by Rader et al. (2013) are compared with the results of normal-hearing (NH) subjects for the following conditions: (1) CI simulation and (2) EAS simulation with low-pass (LP) cutoff frequencies $f_{LP} = 200$ Hz and $f_{LP} = 500$ Hz, i.e., EAS_{200} and EAS_{500}, respectively. NH results for the unprocessed Oldenburg sentence test are shown as a reference. Speech presentation was fixed to 65-dB sound pressure level, and noise level was adjusted using an adaptive procedure.

Table 2, where the aforementioned gain in performance for EAS_{200} to EAS_{500} simulations can be observed. However, compared with NH speech discrimination functions, a general shift of +7.4 dB SNR was found in modulated noise and +10 dB SNR in continuous noise.

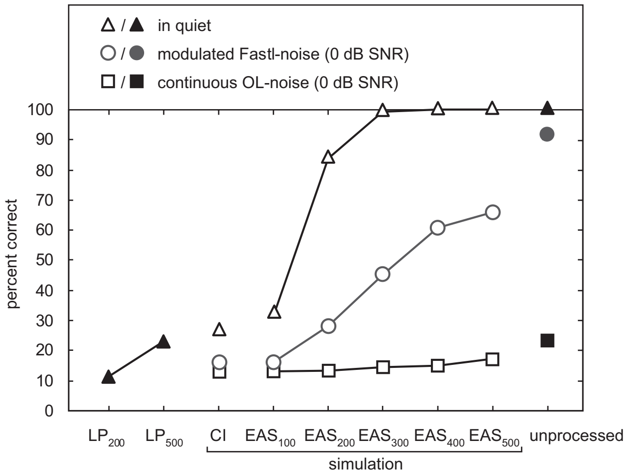

Fig. 5. Speech recognition scores of the automatic speech recognition system in quiet (triangles), in modulated Fastl-noise (circles), and in continuous OL-noise (squares), at a fixed level of 0-dB signal-to-noise ratio (SNR) for the following conditions: (1) low-pass (LP)-filtered speech (closed symbols), (2) cochlear implant (CI) simulation (open symbols), (3) simulation of electric-acoustic stimulation (open symbols), and (4) unprocessed (closed symbols). Cutoff frequencies of LP-filtered speech are $f_{LP} = 200$ and 500 Hz, i.e., LP_{200} and LP_{500}, respectively. For the EAS simulation, cutoff frequencies were $f_{LP} = 100$ to 500 Hz in steps of 100 Hz, i.e., from EAS_{100} up to EAS_{500}.

Comparison With Model Predictions

The synergistic (or “super-additive”) effect of combined stimulation has been more rigorously studied on the basis of perceptual models. Micheyl and Oxenham (2012) summarized recent findings and compared the probability-summation model with two cases of a Gaussian SDT model of cue combination. The two cases were distinguished by the ratio of noise resulting from the combination of information cues (i.e., from LP-filtered and vocoded speech) to any additional noise (e.g., inattention). The late-noise model assumes that this additional noise (referred to as “late” noise) is the only significant source of noise-limiting performance. The independent noise model assumes that the contribution of additional noise is negligible.

To apply these models to the simulation data, probabilities p of a correct response for different SNR levels were estimated from SRT results. To this end, mean SRTs, which correspond to $p = 50\%$, were calculated for EAS_{200} and EAS_{500} simulations. Linear slope approximations (17.1% per dB SNR in modulated and 8% per dB SNR in continuous noise) were then used to estimate p values for the range −6 to +6 dB SNR from the mean SRT, for each of the following conditions: (1) LP_{200} and LP_{500}, respectively, (2) CI simulation, and (3) EAS_{200} and EAS_{500} simulation, respectively. Finally, a numerical solution of the model equation ($m = 8000$; Micheyl & Oxenham 2012) was used to estimate model predictions.

Results for both EAS simulations demonstrated similar trends. To avoid redundancy, Figure 6 shows only model predictions for the EAS_{500} simulation, in modulated Fastl-noise and in continuous OL-noise. As expected from the conclusions of the study by Micheyl and Oxenham (2012), the probability-summation model underpredicted subject performance in all

TABLE 2. SRT of the ASR system in modulated Fastl-noise and continuous OL-noise

	ASR							NH
SRT (dB SNR)	CI	EAS_{100}	EAS_{200}	EAS_{300}	EAS_{400}	EAS_{500}	Unprocessed	Unprocessed
Modulated Fastl-noise	n/a	n/a	11.9	1.4	−2.2	−3.4	−7.5	−14.9
Continuous OL-noise	n/a	n/a	16.8	9.0	7.2	6.3	3.4	−6.6

SRT of the ASR system in modulated Fastl-noise and continuous OL-noise for the following conditions: (1) CI simulation and (2) EAS with LP cutoff frequencies f_{LP} = 100 Hz to f_{LP} = 500 Hz in steps of 100 Hz, i.e., EAS_{100} to EAS_{500}, respectively. The CI and EAS_{100} conditions did not achieve 50% correct speech recognition scores, thus no corresponding SRTs were determined. For comparison, results of the unprocessed Oldenburg sentence test are shown for the ASR system, and for NH subjects in both noise conditions.

ASR, automatic speech recognition; CI, cochlear implant; EAS, electric-acoustic stimulation; LP, low pass; n/a, not applicable; NH, normal hearing; SNR, signal-to-noise ratio; SRT, speech reception threshold.

cases. For the given data, only the late-noise Gaussian SDT model was largely able to predict performance in continuous noise. However, all models significantly underpredicted subject performance in modulated noise. This discrepancy could be attributed to gap listening, which was not accounted for in either of the two SDT model assumptions.

DISCUSSION

Gap Listening Effect in Modulated Noise

The ability to listen to short temporal gaps produced by temporal amplitude fluctuations in modulated noise is a well-known capacity of the auditory system. A “glimpse,” according to Cooke (2006), is a “time–frequency region that contains a reasonably undistorted “view” of local signal properties.” For NH subjects, Li and Loizou (2008) found no difference between speech perception of CI and EAS simulations tested in a continuous speech-shaped noise condition. They used a 5-channel vocoder with added speech LP filtered to 600 Hz to simulate EAS, while adding 3 vocoder channels to substitute LP-filtered speech in the CI simulation. Improved speech recognition for the EAS simulation was only observed in a modulated noise condition (female voice serving as a competing talker).

Accordingly, NH subjects in this study showed significant improvement between the CI and EAS_{200} simulations in modulated Fastl-noise, while such effect was smaller in continuous OL-noise. However, a clear advantage over EAS users was found when NH subjects were tested with the EAS_{500} simulation. EAS users presumably had a greater amount of low-frequency information available to them when compared with NH subjects tested with EAS_{200} simulation (see pure-tone thresholds in the study by Rader et al. 2013). Still, EAS users achieved results similar to NH subjects in the EAS_{200} condition. It is thus evident that NH subjects are superior in taking advantage of the gap listening effect. This was confirmed by the further significant improvement between the EAS_{200} and EAS_{500} simulations, suggesting that the addition of lower harmonics (up to the third harmonic, compare Fig. 2C and D) is beneficial. However, the marked advantage of NH subjects could also be attributed to a shortcoming of simulating residual low-frequency hearing; LP-filtered speech does not incorporate suprathreshold deficits observed in hearing-impaired subjects (Reed et al. 2009, see Different Approaches of EAS Simulation).

Nevertheless, CI and EAS users were generally not able to take advantage of the gap listening effect, conversely having degraded performance when compared with their performance in continuous noise. The frequency resolution in current CI devices is very limited, and the harmonic structure of voiced speech signals is completely distorted. Although the harmonic structure might be preserved in the combined EAS condition because some of the lower harmonics can be detected by residual acoustic hearing (Faulkner et al. 1990; Green et al. 2012), EAS users usually have a steeply sloping hearing impairment. As reported by Rader et al. (2013), the median hearing loss of all ears using EAS was 78 dB HL at 500 Hz. Therefore, strongly degraded frequency and temporal resolution can be expected (Schorn & Zwicker 1990), which may originate from slow rates of recovery from forward masking (Glasberg et al. 1987).

In summary, EAS users have generally better speech perception than do CI users in continuous and modulated noise. The improvement, however, is limited when considering the

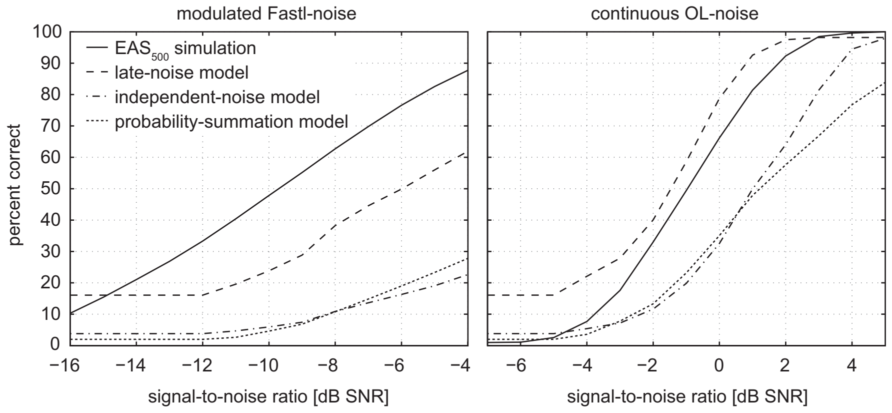

Fig. 6. Measured and predicted performance of normal-hearing subjects (n = 18) tested with simulation of electric-acoustic stimulation (EAS) in modulated Fastl-noise and in continuous OL-noise; low-pass (LP) cutoff frequency was f_{LP} = 500 Hz, i.e., EAS_{500}. Solid lines correspond to direct estimations from speech reception threshold results. Model predictions are shown for the following (*m* = 8000; refer to Micheyl & Oxenham 2012): late-noise model (dashed lines), independent noise model (dash-dotted lines), and probability-summation model (dotted lines). SNR indicates signal-to-noise ratio.

combined-stimulation advantage observed with simulations for NH subjects. It is therefore suggested that residual hearing of EAS users cannot effectively extract acoustic cues.

Impact of Low-Frequency Cutoff in Continuous Noise

In continuous OL-noise, NH subjects tested with CI simulation (i.e., vocoded speech) achieved a median SRT of 5.1 dB SNR, whereas CI users achieved −0.4 dB SNR with significantly better performance distribution ($p < 0.001$). The same effect was true for SRT measured in EAS users compared with the EAS_{200} simulation. Because only very limited training was conducted with NH subjects, these differences would likely decrease with further training (Faulkner 2006). Still, NH subjects tested with the EAS_{500} simulation in continuous noise achieved SRTs that were more comparable with EAS user results, as opposed to SRT results with EAS_{200} simulation. On the one hand, this can be again ascribed to the presumably greater amount of low-frequency hearing available to EAS users compared with the EAS_{200} simulation (see pure-tone thresholds in the study by Rader et al. 2013). On the other hand, the gap listening effect observed for NH subjects showed that the extension of available low-frequency information from 200 to 500 Hz facilitated better speech perception in modulated noise. Thus, in continuous noise, the broader acoustic frequency range may have provided familiar auditory cues to decode speech more effectively and counteract the oddity of the vocoded speech signal (see Supplemental Digital Content 1, http://links.lww.com/EANDH/A190).

Based on the data given by French and Steinberg (1947), an importance function for speech intelligibility was calculated on the basis of the articulation index (ANSI S3.5-1969) as per Fletcher and Galt (1950). In Figure 7, the weight w of the importance function depending on frequency is adopted from Terhardt (1998). Interestingly, w shows a strong increase starting with $w = 0$ at $f = 100$ Hz, reaching a plateau at 700 Hz. Between 200 and 500 Hz, the importance function increases steeply from $w = 0.1$ to $w = 0.36$, whereas between 500 Hz and 1 kHz, there is little change. In other words, the extension of available low-frequency information below 500 Hz can certainly have a positive impact on speech perception. Our results demonstrated this for NH subjects, where they had significant improvement in performance between the EAS_{200} and EAS_{500} simulations in continuous noise. Still, importance functions can vary between different speech materials, especially when used in predictions for hearing-impaired subjects (Pavlovic 1986). While the articulation index successor, the Speech Intelligibility Index (ANSI S3.5-1997), attempted to correct this, individual importance functions have been shown to provide more reliable results (Whitmal & DeRoy 2011).

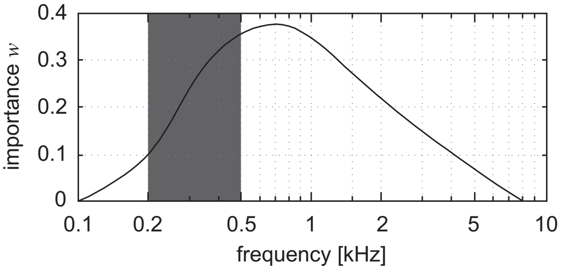

Fig. 7. Importance w of spectral frequencies for the identification of senseless syllables (adopted from Terhardt 1998). Calculation based on articulation index as per French and Steinberg (1947). Shaded area indicates relevant frequency range from 200 to 500 Hz as in the Oldenburg sentence test using electric-acoustic simulations for normal-hearing subjects.

Impact of Low-Frequency Cutoff on ASR

Regarding the synergistic effect of combined EAS in quiet, results obtained with the ASR system demonstrated a similar behavior. As can be seen in Figure 5, LP-filtered speech with cutoff frequency $f_{\text{LP}} = 500$ Hz (i.e., LP_{500}) yielded a speech recognition score of 22.3%, while vocoded speech (i.e., CI simulation) yielded a recognition score of 26.3%. The combined signal consisting of LP-filtered and vocoded speech (i.e., EAS_{500}) reached nearly perfect recognition (99.9%) in quiet. Even with lower cutoff frequencies, the "super-additive" effect was observable. However, this effect broke down for the lowest cutoff frequency tested (i.e., EAS_{100}), where a score of only 32.5% was reached. This is close to the recognition score obtained for the CI simulation without any complementary low-frequency acoustic information, which can be due to the fundamental frequency range generally lying above 100 Hz for OLSA (Fig. 2).

The preprocessing of the ASR system applies a filter bank with a mel-scale distribution of band-pass center frequencies and bandwidths to code input signals into feature vectors. The mel-scale is a frequency-to-cochlear-position map where equal pure-tone pitch differences correspond to equal cochlear distances, as per Stevens and Volkmann (1940). This affects a setting of narrower bandwidths for lower center frequencies and broader filters for higher frequencies, resulting in potentially higher frequency selectivity for signals with lower frequency content, thereby mimicking the characteristics of the human auditory system. Consequently, it was expected that the ASR system and NH subjects exhibit similar behavior, whereas EAS users are known to have reduced frequency selectivity at low frequencies (e.g., Moore et al. 1997).

Applied Model Based on ASR

Our customized ASR model scored 99.9% correct on word recognition with the closed-set OLSA speech material, justifying our ASR model approach to study the effect of LP-filtered and vocoded speech on speech recognition stability. While the ASR system achieved better SRTs in modulated noise than in continuous noise, as observed in NH subjects, a general shift of the speech discrimination functions was found. This shift was +7.4 dB SNR in modulated noise and +10 dB SNR in continuous noise, which reflects the poor performance of current ASR implementations for speech in noise. A possible reason for these shifts could be degraded onset and offset detection due to competing noise. Supplemental state transitions were added to monophone HMMs to avoid erroneous transitions from impulsive noises. The monophone HMMs were then tied into three-state left-to-right triphone HMMs (Young et al. 2006). Broadband noise, however, still obscured word boundaries that the model struggled to separate context-dependent triphones.

The slope of the speech discrimination function obtained for the ASR system in continuous noise was flatter than the slope measured for NH subjects (8.4% versus 17.1% per dB SNR). Because the continuous noise consisted of the same (averaged) signal as the target speech, a performance bias was potentially introduced. Namely, HMM training was performed using OLSA speech material, while the continuous OL-noise was generated by an averaging process of 30 time-shifted OLSA test sentences (see Noise Characteristics). A better investigation of the OL-noise composition in short time segments, which are relevant for coding the speech feature vectors, could result in better model composition.

The ASR system, despite a few drawbacks, has proven to be a useful tool for the applied model approach, mimicking subject performance for a given psychophysical task. It is feasible that more robust and general models can be developed, allowing efficient preliminary assessments of perceptual model hypotheses.

Prediction Model Based on SDT

As thoroughly explained in the study by Micheyl and Oxenham (2012), the Gaussian SDT model provides a solution to an *m*-alternative forced-choice task. The subject's task can be thus seen as identifying a speech item (e.g., a word) drawn from a set of *m* items. The numerically solved equation for the model comparison calculates the probability of a correct response as a function of *m* and *d'*, where *d'* equals the ratio of the mean to standard deviation of a decision variable. Given that OLSA is a closed-set task with 50 items, we first applied the model for *m* = 50. The results, however, underpredicted user performance for all conditions. Green and Birdsall (1958) suggested that when the stimulus set is large and not known to the subject beforehand, then performance is determined by the number of all possible responses and not limited by the number of alternatives presented. In this case, possible responses are limited only by the size of the subject's active vocabulary, which Müsch and Buus (2001) estimated to be about 8000 words. While this estimation was not made for the German language, and although our results were better fitted for other *m* values (e.g., *m* = 5000), the overall model prediction remained comparable. Consequently, we preferred assessing our data with the previously used parameter *m* = 8000. Within the given framework, Gaussian SDT-based models were largely able to account for the "synergistic effect," which was shown in continuous noise with the late-noise model, but not for the independent noise model. Furthermore, the gap listening effect was not accounted for in the model assumptions and thus all investigated models fell short of predicting the combined-stimulation advantage in modulated noise. While the SDT model falls in the category of prelabeling models, which assume that subjects combine information across channels before making a decision, the "clean" information channels inherently provided by gap listening were not mapped in the current model assumptions. As previously suggested (Seldran et al. 2011), it might be necessary to include across-channel interactions in SDT models to resolve the observed discrepancies.

In conclusion, it is plausible that predictions based on SDT can explain the combined-stimulation advantage without the need to assume any synergistic effects. The simplified model assumptions, however, are not yet able to account for the performance gain in modulated noise, which allowed gap listening in NH subjects tested with EAS simulations.

Different Approaches of EAS Simulation

The psychophysical mechanism of the combined-stimulation advantage is often investigated by means of simulation studies; these evaluate speech intelligibility in noise for NH subjects tested with simulations of CI and EAS hearing (see Introduction). Consequently, such investigations depend on the appropriate simulation of electric hearing, on the simulation of residual acoustic hearing, and on the choice of competing noise.

The simulation of CI hearing commonly uses the temporal envelope of speech to modulate narrow frequency bands corresponding to each CI channel. This can be achieved using sinusoidal vocoders or noise-band vocoders (Shannon 1995). While sine carriers form a rather simplified model of electric hearing, effects of intrinsic modulations in noise carriers are still debated regarding their use in research on speech perception. Other approaches constructed broadband signals with low intrinsic modulations after auditory filtering (Hilkhuysen & Macherey 2014). Using such signal designs might prove more appropriate to simulate CI hearing for future studies.

Furthermore, presenting LP-filtered speech to simulate residual acoustic hearing does not incorporate suprathreshold deficits observed in hearing-impaired subjects (Reed et al. 2009), for example, loudness recruitment or the degraded ability to use temporal fine-structure information. Thus, NH subjects are expected to take the advantage of speech cues in LP-filtered speech more effectively than EAS users. It is suggested that using simulations that mimic suprathreshold deficits in hearing-impaired subjects could provide better simulations of residual low-frequency hearing. Different modes of presentation can also influence simulation results; in the present study, NH subjects were presented with diotic EAS simulations and compared with bimodal EAS users. Depending on audiometric configuration and amplification characteristics, aided acoustic hearing in the contralateral ear can interfere with the information provided by EAS. However, all bimodal EAS users discussed in this study showed no negative interference because their bimodal speech perception scores in noise were higher compared with results of the aided acoustic hearing alone.

Finally, the choice of competing noise can influence the observed combined-stimulation advantage; Turner et al. (2004) clearly demonstrated an advantage of combined EAS when tested in competing speech but showed nonsignificant advantage in continuous noise. However, in our previous study (Rader et al. 2013), we used the contrast of pseudocontinuous and amplitude-modulated noise to examine aspects of gap listening and bilateral interaction without introducing informational masking as per competing speech. In the present study, a significant advantage was shown for NH subjects in both noise conditions when comparing the CI simulation with either EAS simulations. Between the two masker conditions, improved performance in modulated noise substantiated the gap listening effect in the simulation and model comparison.

CONCLUSIONS

The results of the present study demonstrated the advantage of combined EAS; NH subject performance significantly improved when vocoded speech simulating CI hearing was complemented with LP-filtered speech. Increasing the LP cutoff frequency from 200 to 500 Hz, and thereby increasing the amount of low-frequency acoustic information, significantly improved SRTs in pseudocontinuous and in amplitude-modulated noise. NH subjects were able to take advantage of the gap listening effect when tested with EAS simulations. In contrast, CI and EAS user performance was consistently degraded in modulated noise compared with their performance in continuous noise.

A model approach based on a customized ASR system yielded better SRTs in modulated noise than in continuous noise. Thus, the gap listening effect observed in NH subjects was successfully imitated by the ASR system. Results obtained with this model also demonstrated the positive effect of low-frequency spectral information, which supplemented vocoded speech of a male talker. Our results suggest that speech recognition stability significantly improves for LP cutoff frequencies ≥300 Hz.

While Gaussian SDT-based models were largely able to predict the combined-stimulation advantage in continuous noise without assuming any synergistic effects, model amendments are required to explain the performance gain generated by the gap listening effect in modulated noise.

ACKNOWLEDGMENTS

The authors thank Daniel Menzel, Florian Völk, and Jakob Putner of the Technical University in Munich for their support with experimental data acquisition; Felix Weninger for providing the German language corpus for the HTK Toolkit; and Michael Todd for improvements to our English language. The authors especially thank all test subjects for their participation. The study was approved by the local ethical review board (University of Frankfurt, 168/07).

The study was supported by a grant from the German Research Foundation (DFG BA 2085/3-1).

Portions of this article were presented at the 2011 Conference on Implantable Auditory Prostheses (CIAP) July 24–29, 2011, at Asilomar, Pacific Grove, and at the 2014 Speech in Noise Workshop (SpiN) January 9–10, 2014, in Marseille, France.

The authors declare no other conflict of interest.

Address for correspondence: Tobias Rader, Universitätsklinikum Frankfurt, KHNO/Audiologische Akustik, Theodor-Stern Kai 7, 60590 Frankfurt am Main, Germany. E-mail: tobias.rader@kgu.de

Received August 29, 2014; accepted March 26, 2015.

REFERENCES

ANSI S3.5-1969. (1969). *American National Standard Methods for the Calculation of the Articulation Index*. New York, NY: American National Standards Institute.

ANSI S3.5-1997. (1997). *American National Standard Methods for the Calculation of the Speech Intelligibility Index*. New York, NY: American National Standards Institute.

Baumann, U. (1995). Ein Verfahren zur Erkennung und Trennung multipler akustischer Objekte [A procedure for identification and segregation of multiple auditory objects]. Doctoral dissertation. Technische Universität München, Munich, Germany.

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J Acoust Soc Am*, *111*, 2801–2810.

Brown, C. A., & Bacon, S. P. (2010). Fundamental frequency and speech intelligibility in background noise. *Hear Res*, *266*, 52–59.

Ching, T. Y., van Wanrooy, E., Dillon, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends Amplif*, *11*, 161–192.

Cong-Thanh Do, Pastor, D., Goalic, A. (2010). On the recognition of cochlear implant-like spectrally reduced speech with MFCC and HMM-based ASR. *IEEE T Audio Speech*, *18*, 1065–1068.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *J Acoust Soc Am*, *119*, 1562–1573.

Cullington, H. E., & Zeng, F. G. (2010). Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. *Ear Hear*, *31*, 70–73.

Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *Int J Audiol*, *49*, 912–919.

Fastl, H. (1987). Ein Störgeräusch für die Sprachaudiometrie [A background noise for speech audiometry]. *Audiologische Akustik*, *26*, 2–13.

Fastl, H. & Zwicker, E. (2007). *Psychoacoustics: Facts and Models* (3rd ed., Vol. 22). Berlin, New York: Springer.

Faulkner, A. (2006). Adaptation to distorted frequency-to-place maps: Implications of simulations in normal listeners for cochlear implants and electroacoustic stimulation. *Audiol Neurootol*, *11*(Suppl 1), 21–26.

Faulkner, A., Rosen, S., Moore, B. C. (1990). Residual frequency selectivity in the profoundly hearing-impaired listener. *Br J Audiol*, *24*, 381–392.

Fletcher, H. & Galt, R. (1950). The perception of speech and its relation to telephony. *J Acoust Soc Am*, *22*, 89–150.

French, N. & Steinberg, J. (1947). Factors governing the intelligibility of speech sounds. *J Acoust Soc Am*, *19*, 90–119.

Gantz, B. J., & Turner, C. W. (2003). Combining acoustic and electrical hearing. *Laryngoscope*, *113*, 1726–1730.

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

Gifford, R. H., & Dorman, M. F. (2012). The psychophysics of low-frequency acoustic hearing in electric and acoustic stimulation (EAS) and bimodal patients. *J Hear Sci*, *2*, 33–44.

Gifford, R. H., Dorman, M. F., McKarns, S. A., et al. (2007). Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. *J Speech Lang Hear Res*, *50*, 835–843.

Glasberg, B. R., Moore, B. C., Bacon, S. P. (1987). Gap detection and masking in hearing-impaired and normal-hearing subjects. *J Acoust Soc Am*, *81*, 1546–1556.

Green, D. & Birdsall, T. (1958). The effect of vocabulary size on articulation score: Technical Report No. 81. Ann Arbor, MI: University of Michigan.

Green, T., Faulkner, A., Rosen, S. (2012). Frequency selectivity of contralateral residual acoustic hearing in bimodal cochlear implant users, and limitations on the ability to match the pitch of electric and acoustic stimuli. *Int J Audiol*, *51*, 389–398.

Heinbach, W. (1988). Aurally adequate signal representation—The Part-Tone-Time-Pattern. *Acustica*, *67*, 113–121.

Helbig, S., Baumann, U., Hey, C., et al. (2011). Hearing preservation after complete cochlear coverage in cochlear implantation with the free-fitting FLEXSOFT electrode carrier. *Otol Neurotol*, *32*, 973–979.

Hilkhuysen, G., & Macherey, O. (2014). Optimizing pulse-spreading harmonic complexes to minimize intrinsic modulations after auditory filtering. *J Acoust Soc Am*, *136*, 1281–1294.

Incerti, P. V., Ching, T. Y., Cowan, R. (2013). A systematic review of electric-acoustic stimulation: Device fitting ranges, outcomes, and clinical fitting practices. *Trends Amplif*, *17*, 3–26.

International Organization for Standardization (ISO) 7029:2000. (2001). *Acoustics—Statistical Distribution of Hearing Thresholds as a Function of Age*. Geneva: International Organization for Standardization.

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

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

Micheyl, C., & Oxenham, A. J. (2012). Comparing models of the combined-stimulation advantage for speech recognition. *J Acoust Soc Am*, *131*, 3970–3980.

Moore, B. C., Vickers, D. A., Glasberg, B. R., et al. (1997). Comparison of real and simulated hearing impairment in subjects with unilateral and bilateral cochlear hearing loss. *Br J Audiol*, *31*, 227–245.

Müsch, H., & Buus, S. (2001). Using statistical decision theory to predict speech intelligibility. I. Model structure. *J Acoust Soc Am*, *109*, 2896–2909.

Pavlovic, C. V., Studebaker, G. A., Sherbecoe, R. L. (1986). An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. *J Acoust Soc Am*, *80*, 50–57.

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

Rader, T., Fastl, H., Baumann, U. (2013). Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field. *Ear Hear*, *34*, 324–332.

Reed, C. M., Braida, L. D., Zurek, P. M. (2009). Review article: Review of the literature on temporal resolution in listeners with cochlear hearing impairment: A critical assessment of the role of suprathreshold deficits. *Trends Amplif*, *13*, 4–43.

Schorn, K. & Zwicker, E. (1990). Frequency selectivity and temporal resolution in patients with various inner ear disorders. *Audiology*, *29*, 8–20.

Seeber, B., Fastl, H., Koči, V. (2003). Ein PC-basiertes Békésy-Audiometer mit Bark-Skalierung [PC-based Békésy audiometer with Bark scaling]. In M. Vorländer (Ed.), *Tagungsband DAGA* (pp. 614–615). Oldenburg, Germany: Isensee.

Seldran, F., Micheyl, C., Truy, E., et al. (2011). A model-based analysis of the "combined-stimulation advantage." *Hear Res*, *282*, 252–264.

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Stevens, S. & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *Am J Psychol*, *53*, 329–353.

Terhardt, E. (1998). *Akustische Kommunikation: Grundlagen mit Hörbeispielen [Acoustic Communication: Fundamentals With Audio Samples]*. Berlin, Germany: Springer.

Terhardt, E. (1985). Fourier transformation of time signals: Conceptual revision. *Acustica*, *57*, 242–256.

Turner, C. W., Gantz, B. J., Vidal, C., et al. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *J Acoust Soc Am*, *115*, 1729–1735.

Vertanen, K. (2006). *Baseline WSJ Acoustic Models for HTK and Sphinx: Training Recipes and Recognition Experiments*. Cambridge, UK: Cambridge University Press.

von Ilberg, C. A., Baumann, U., Kiefer, J., et al. (2011). Electric-acoustic stimulation of the auditory system: A review of the first decade. *Audiol Neurootol*, *16*(Suppl 2), 1–30.

von Ilberg, C., Kiefer, J., Tillein, J., et al. (1999). Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss. *ORL J Otorhinolaryngol Relat Spec*, *61*, 334–340.

Wagener, K., Brand, T., Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil III: Evaluation des Oldenburger Satztests [Development and evaluation of a German sentence test. Part III: Evaluation of the Oldenburg sentence test]. *Z Audiol*, *38*, 86–95.

Wagener, K., Brand, T., Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil II: Optimierung des Oldenburger Satztests [Development and evaluation of a German sentence test. Part II: Optimization of the Oldenburg sentence test]. *Z Audiol*, *38*, 44–56.

Wagener, K., Kühnel, V., Kollmeier, B. (1999c). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. Teil I: Design des Oldenburger Satztests [Development and evaluation of a German sentence test. Part I: Design of the Oldenburg sentence test]. *Z Audiol*, *38*, 4–15.

Wells, C. J. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, R. Winski (Eds.), *Handbook of Standards and Resources for Spoken Language Systems (Part IV, section B)*. Berlin, New York: Mouton de Gruyter.

Weninger, F., Schuller, B., Eyben, F. et al. (2010). A Broadcast News Corpus for Evaluation and Tuning of German LVCSR Systems. Munich, Germany: Technische Universität München.

Whitmal, N. A. III, & DeRoy, K. (2011). Adaptive bandwidth measurements of importance functions for speech intelligibility prediction. *J Acoust Soc Am*, *130*, 4032–4043.

Young, S., Evermann, G., Gales, M. et al. (2006). *The HTK Book (for HTK version 3.4)*. Cambridge, UK: Cambridge University Press.

Zokoll, M. A., Hochmuth, S., Warzybok, A., et al. (2013). Speech-in-noise tests for multilingual hearing screening and diagnostics1. *Am J Audiol*, *22*, 175–178.

Zwicker, E. & Heinz, W. (1955). Zur Häufigkeitsverteilung der menschlichen Hörschwelle [Contribution to the distribution of the human hearing threshold]. *Acustica*, *5*(Suppl 1), 75–80.