Audibility emphasis of low-level sounds improves consonant identification while preserving vowel identification for cochlear implant users

Raymond L. Goldsworthy [1], Susan R.S. Bissmeyer [1,2], Jayaganesh Swaminathan [3,4]

[1] Auditory Research Center, Caruso Department of Otolaryngology, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America
[2] Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, California, United States of America.
[3] Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts, United States of America
[4] Department of Audiology, University of the Pacific, San Francisco, California, United States of America

ARTICLE INFO

ABSTRACT

Consonant perception is challenging for listeners with hearing loss, and transmission of speech over communication channels further deteriorates the acoustics of consonants. Part of the challenge arises from the short-term low energy spectro-temporal profile of consonants (for example, relative to vowels). We hypothesized that an audibility enhancement approach aimed at boosting the energy of low-level sounds would improve identification of consonants without diminishing vowel identification. We tested this hypothesis with 11 cochlear implant users, who completed an online listening experiment remotely using the media device and implant settings that they most commonly use when making video calls. Loudness growth and detection thresholds were measured for pure tone stimuli to characterize the relative loudness of test conditions. Consonant and vowel identification were measured in quiet and in speech-shaped noise for progressively difficult signal-to-noise ratios (+12, +6, 0, -6 dB SNR). These conditions were tested with and without an audibility-emphasis algorithm designed to enhance consonant identification at the source. The results show that the algorithm improves consonant identification in noise for cochlear implant users without diminishing vowel identification. We conclude that low-level emphasis of audio can improve speech recognition for cochlear implant users in the case of video calls or other telecommunications where the target speech can be preprocessed separately from environmental noise.

INTRODUCTION

Cochlear implants have been remarkably successful for restoring hearing to people with hearing loss. Today, more than half-a-million people have received cochlear implants, and recipients generally achieve high levels of speech comprehension without needing to rely on visual speech-reading cues. While the typical recipient can reach high levels of speech comprehension in quiet after a year of rehabilitation with their device, speech comprehension in noise remains poor compared to their normal-hearing peers (Firszt et al., 2004; Niparko et al., 2010). Further, there is a remarkably broad range of outcomes across cochlear implant recipients with a large portion not obtaining open-set speech comprehension even in quiet listening situations (Wilson and Dorman 2008). Compared to normal-hearing listeners, consonant identification for cochlear implant users is degraded more than vowel identification when listening in background noise (Goldsworthy et al., 2013; Goldsworthy 2015). This finding motivates the present study, which examines audibility emphasis of low-level sounds as a method to enhance consonant perception.

That consonant identification is degraded more than vowel identification when testing people with hearing loss in background noise can be understood by considering physiology of normal hearing. It is well-established that active mechanisms in healthy hearing contribute to amplification of relatively low-level sounds (Smith 1979; Delgutte and Kiang 1984). These active mechanisms degrade with hearing loss making consonants, which are relatively low-level transient sounds compared to vowels, difficult to perceive. In addition to audibility concerns with consonants, the degradation of spectral resolution and modulation detection likely contribute to the relative deficit of consonant compared to vowel identification in the presence of background noise (Dreschler and Plomp 1985; Hochberg et al., 1992; Drullman et al., 1994; Friesen et al., 2001; Bhattacharya and Zeng 2007).

Cochlear implants and hearing aids attempt to correct for audibility by specifying a loudness growth function that allows soft sounds to be heard while providing loudness discrimination for medium to loud sounds (Zeng et al., 2002, 2005). Much effort has been given to automatic gain control for hearing devices with emphasis on optimal parameters for providing good speech recognition (Blamey 2005; Baker and Sarpeshkar 2006; Won et al., 2011). Nevertheless, evidence suggests that consonant identification remains relatively poor compared to vowel

identification when testing in background noise. Specifically, while cochlear implant users perform worse on both consonant and vowel identification measures compared to their normal-hearing peers, the relative loss is greater for consonant identification (Goldsworthy et al., 2013; Goldsworthy 2015).

Recognizing the relevance of this problem for cochlear implant outcomes, work has been conducted to determine if audibility and/or transient emphasis could improve speech comprehension in background noise for cochlear implant users. The transient emphasis spectral maximum approach applies additional gain to time-frequency components that exhibit rapid changes (Vandali and Clark 2010). This approach improved speech reception in multiple-talker babble noise, including in combination with spectral expansion (Bhattacharya et al., 2011). Specifically, sentence recognition improved with both spectral expansion alone and spectral expansion combined with the Transient Emphasis Spectral Maxima (TESM) approach. In an earlier study, it was shown that the benefit derived from that approach appears to be driven by audibility of consonants since no benefit was found when the input speech was at a relatively high 75 dB sound pressure level (Holden et al., 2005).

Following this work, several studies focused on onset enhancement particularly, rather than audibility or transient emphasis generally, to improve speech comprehension for cochlear implant users (Choi et al., 2008; Koning and Wouters 2016). An envelope enhancement (EE) algorithm was shown to improve the intelligibility of speech in noise for normal-hearing participants listening to acoustic "vocoder" simulations of cochlear implant sound processing (Koning and Wouters 2012). In these studies, the enhancement was applied to low-level transients while controlling for the broadband SNR. The measured benefit in speech reception threshold (SRT) was 2.5 dB in speech-shaped noise and 2.6 dB in competing speech. For both results, the algorithm had access to the clean target speech for pre-emphasis.

These studies motivate careful examination of consonant and vowel identification in the context of audibility of low-level sounds. The approach described here operates on the broadband signal and can be easily implemented in real-time signal processing. Consonant enhancement is a notoriously challenging problem, but in situations where the target speech is known, such as in telecommunications, preprocessing of the source material can lead to significant benefits under several realistic listening conditions. We hypothesized that audibility emphasis of low-level sounds at the source would improve identification of consonants without diminishing vowel identification. To test this hypothesis, consonant identification was measured in cochlear implant users in quiet and in speech-shaped noise at several signal-to-noise ratios (SNRs). Audibility emphasis was applied to the speech material prior to mixing with noise to emulate the situation that people encounter when speaking on the phone in a noisy environment, specifically, where the target speech is available for pre-processing. The same emphasis algorithm was applied to speech materials designed for vowel identification to determine if enhancing low-level sounds would cause any deficit in vowel identification. The results indicate that audibility emphasis of low-level sounds provides a large and significant benefit for consonant identification while not significantly affecting vowel identification. Consequently, we conclude that this approach for audibility emphasis of low-level sounds is promising as a method to improve speech comprehension in noisy situations, at least for the telecommunication situation where the target speech is known.

METHODS

Participants

Eleven adult cochlear implant users took part in this study. Five of the participants were bilateral of whom two chose to test with each ear separately, with the first ear randomly selected. Participants provided informed consent and were paid for their participation. This study was approved by the University of Southern California Institutional Review Board. Participant information is provided in Table 1.

Overview

Adult cochlear implant users took part in an experiment designed to test whether audibility emphasis of low-level sounds can improve consonant identification in background noise without diminishing vowel identification. All testing was conducted remotely with participants connecting to computer audio from their homes. Loudness scaling and pure-tone detection thresholds were first measured to determine loudness levels and to reference these levels to their detection thresholds. Following loudness scaling and detection thresholds, participants completed consonant and vowel identification procedures with original and enhanced stimuli at progressively more difficult SNRs. The ordering of original and enhanced phoneme presentation was randomized, and all participants were tested on consonant and vowel identification including conditions without noise and with speech-spectrum noise with +12, +6, 0, and -6 dB SNRs. Specifically, subjects completed two repetitions of the original/enhanced conditions in random order in quiet followed by two repetitions of the conditions in random order at +12 dB SNR. These conditions in quiet and with +12 dB SNR were treated as familiarization procedures with subjects completing two runs of each. Following those conditions, subjects completed testing in order of increasingly difficulty, with identification measured at +6, 0, and -6 dB SNRs tested with three runs of the original/enhanced conditions in random order. Each run consisted of 20 trials for consonant identification (one repetition of each consonant) and 24 trials for vowel identification (two repetitions of each vowel) runs of each and analyses focus on these conditions, which followed the familiarization procedures. Feedback was provided during all procedures by displaying a green checkmark after correct answers and a red "X" after mistakes.

Loudness scaling

Categorical loudness scaling was measured for a 1 kHz pure tone. An application interface was provided that prompted participants to adjust the tone level to four loudness levels described as "Soft", "Medium Soft", "Medium", and "Medium Loud". Participants could adjust gain by 2 dB increments or decrements. After adjusting loudness levels, participants pressed an "Okay" button and the level specified as "Medium" was used for subsequent phoneme identification procedures. We chose 1 kHz as the comparison frequency because of its central position in predicting speech intelligibility (Steeneken and Houtgast 2002); though we note that the current international standard recommends an extended set of frequencies from 500 to 4000 Hz to characterize loudness (ISO 16832).

Pure tone detection thresholds

Pure tone detection thresholds were measured for 500, 1000, 2000, and 4000 Hz tones. Tones were 400 ms sinusoids with 20 ms raised-cosine attack and release ramps. The initial level of the sinusoid was set to the "Medium" level specified in the loudness scale and a three-alternative forced-choice (oddball) procedure was used to measure detection thresholds. Following correct responses, the presentation level was decreased by 2 dB and, following wrong responses, increased by 6 dB. A measurement run continued until 3 mistakes were made with all mistakes counted including sequential mistakes. The average of the last 4 reversals was taken as the threshold estimate.

Phoneme Identification

Consonant and vowel identification were measured for originally recorded materials and for materials processed with an audibility-emphasis algorithm. Consonants were drawn from speech samples collected by Shannon et al., 1999 for five male and five female talkers and consisted of twenty phonemes /b tʃ d f g dʒ k l m n p r s ʃ t ð v w y z/, presented in /a/–C–/a/ context (aba, acha, ada, afa, aga, aja, aka, ala, ama, ana, apa, ara, asa, asha, ata, atha, ava, awa, aya, aza). Vowels were drawn from speech samples collected by Hillenbrand et al., 1995,
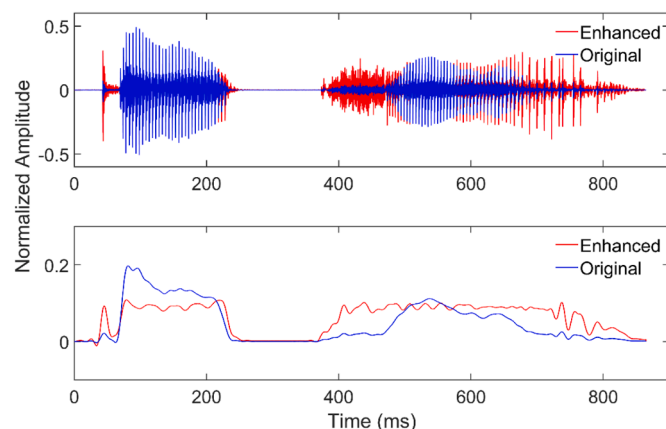
TABLE 1
Participant information.

Subject	Gender	Ear Tested	Manufacturer	Etiology	Age at Onset of Hearing Loss	Age at Deafness	Age at Implantation	Age at Time of Testing	Method of Streaming
C1	M	Both	Cochlear	Meniere's	39	L:46 R:39	L:46 R:43	47	Mini Mic
C2	F	Both	Cochlear	Unknown	15	22	L:23 R:27	34	Mini Mic2
C3	F	Both Together	Cochlear	Progressive Nerve Loss	40	53	L:54 R: 58	72	Cochlear Binaural Cable
C4	M	Left	Advanced Bionics	Autoimmune Disease	54	55	56	56	Bluetooth/Compilot
C5	M	Right	Cochlear	Noise Induced	40	50	70	80	Juster Multimedia Speaker SP-689
C6	F	Right	Cochlear	Sudden Nerve Loss	40	50	69	70	Free Field through iPad Speakers
C7	M	Both Together	Cochlear	Unknown	Birth	10	L:71 R:70	73	Mini Mic2
C8	F	Right	Advanced Bionics	Sudden Nerve Loss	55	55	56	57	AB Bluetooth Streaming
C9	M	Right	Cochlear	Unknown	40	46	48	48	Mini Mic
C10	M	Right	Med-El	Mumps Disease	14	14	56	58	I-loop streaming
C11	F	Both Together	Advanced Bionics	Ototoxic Medicine	4	10	L:64 R:43	70	Free Field through MacAir Book Speakers

for five male and five female talkers and consisted of ten monophthongs (/i ɪ ɛ æ u ʊ ɑ ɔ ʌ ɝ/) and two diphthongs (/əʊ eɪ/), presented in /h/-V-/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, hayed). Listeners responded using a graphical user interface with twenty (consonants) or twelve (vowels) alternatives with the appropriately labeled phonemes. All phonemes were presented at the same RMS value as a 1 kHz pure tone that was specified as "Medium" loud by each subject.

The audibility enhancement algorithm is based on the energy-equalization approach which acts to normalize rapidly fluctuating short-term signal energy to be equal to the long-term average signal energy (Deslonge et al., 2017). This technique acts to first separate the signal into short time-frame segments (e.g., 5 ms), compute the signal energy in each short time-frame segment, increase the level of the short-term energy in segments for which the energy is below a set threshold and then combine the short-term segments to create a long signal with the same duration as the original unprocessed signal. The processed signal is normalized such that the overall level of the enhanced signal is equal to the input signal. This approach works naturally for enhancing the energy of the consonants relative to vowels (thereby adjusting the consonant-vowel energy ratio within an utterance)(Swaminathan et al., 2020).

This approach also enhances the onset slope at the regions of the consonants as illustrated in Fig. 1, which shows the original /VCV/ waveform 'APA' (upper panel in blue) and consonant enhanced in red (upper panel in red). The broadband low pass filtered/smoothed envelope is shown in the lower panel. The energy and the onset broadband envelope slope at the consonant /p/ are enhanced compared to the unprocessed signal. Note that both signals are normalized to have the same overall level.

Fig. 2 illustrates the effect of the enhancement algorithm on the electrical stimulation patterns used with cochlear implant stimulation. The electrical stimulation patterns (or "electrodograms") were generating by processing original and enhanced tokens through the Nucleus MATLAB Toolbox made available by Cochlear Corporation with default parameters. The enhancement algorithm results in stronger electrical stimulation in the region near 400 ms associated with the consonant sound.

For the noise conditions, speech-spectrum noise was generated by filtering random noise drawn from a uniform distribution through a speech-spectrum shaping filter. This filter was generated by estimating the power spectral density of the corresponding speech corpus (i.e., consonants or vowels) using Welch's periodogram method and converting this density to an 8th-order IIR filter using Prony's method (Oppenheim 1978). A 20 second sample of noise was generated and played continuously throughout the phoneme identification procedures
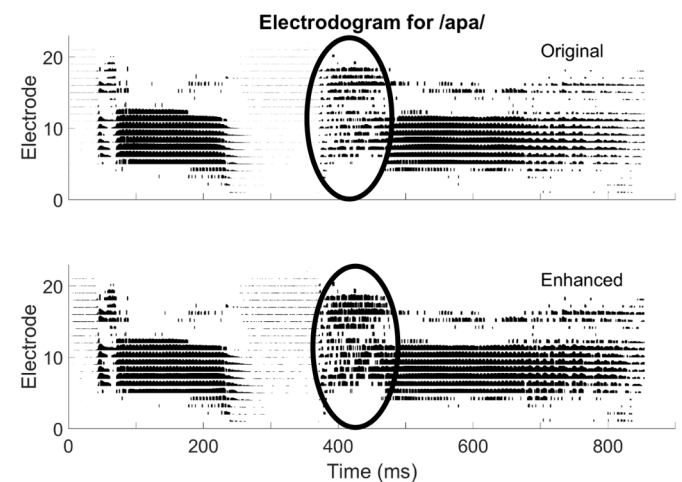

Fig. 1. Representative broadband audio file and amplitude envelopes for original and enhanced phonemes (a token /apa/ from the consonant database is used in this example). The effect of enhancement on the consonant is strong near 400 ms.

Fig. 2. Representative electrical stimulation patterns for original and enhanced phonemes (as in Fig. 1). The ovals overlaid near 400 ms display the stimulus region associated with the /p/ sound, which clearly has a stronger electrical representation compared to the original sound.

(except for the quiet condition). Brief 20 ms attack and release ramps were applied at the beginning and end of the noise sample to avoid snap, crackle, and pop artifacts between loops. The SNR was specified based on the root mean square value of the phoneme and noise samples. For a given SNR, the phoneme and noise samples were combined and scaled such that the total output power was set equal to the subject-specified "Medium" loudness level for a 1 kHz pure tone as described in the loudness scaling section. Doing so provides a consistent reference, but it is unclear to what extent scaling the phoneme materials in this manner produces equally loud percepts.

Consonant and vowel identification were measured using a web application with participants using an application interface with twenty consonant or twelve vowel alternatives with appropriately labeled phonemes. Feedback was provided in the form of the pressed button flashing a green checkmark for correct answers or a red "X" for wrong answers. A run of the consonant identification measure consisted of 20 trials corresponding to one presentation of each consonant. A run of the vowel identification measure consisted of 24 trials corresponding to two presentations of each vowel.

Participants were tested at progressively more difficult SNR conditions. Participants were first tested on two measurement runs of consonant and vowel identification in quiet for both the original and enhanced materials. Participants were then tested on two runs for both consonants and vowels at +12 dB for both the original and enhanced materials. These initial runs were treated as familiarization exercises with and analyses focus on the +6, 0, and -6 dB SNR conditions. Following the +12 dB SNR condition, participants were then tested using three measurement runs for both consonant and vowels and for both original and enhanced materials. The ordering of original and enhanced material presentation was randomized across runs. While the ordering of SNR conditions from more to less favorable SNR conditions may provide a slight familiarization advantage (which also might be offset by fatigue), we note that our primary hypothesis concerns the effect of the transient-emphasis algorithm, which was randomized across conditions. A permalink for this experiment can be found at: https://www.teamhearing.org/79. Upon entering the site, click the "Homework" button to start the experiment.

Data Analysis

The working hypothesis is that audibility emphasis of low-level sounds can improve consonant identification in background noise without significantly affecting vowel identification. The experimental design was full factorial with repeated measures. Measures were converted from percent correct to rationalized arcsine units. Separate analyses were conducted on the measures for consonant and vowel identification. The analyses were repeated-measures analysis of variance with processing condition (original or enhanced) and SNR (+6, 0, and -6 dB) as within-subject factors. Planned multiple comparisons were made comparing identification scores for original and enhanced materials at each SNR. Cohen's d was calculated as a measure of effect size (Cohen 1992).

RESULTS

Loudness scaling and detection thresholds

Fig. 3 shows categorical loudness settings and detection thresholds, which were used to provide a relative anchor as to presentation level of the test materials. Loudness ratings reflect aspects of cochlear implant programming, particularly the loudness growth function. The phoneme identification procedures were conducted with stimuli normalized to have the same output root-mean-square level as a 1 kHz pure tone presented at the subjectively defined "Medium" level. The average value of this medium loudness level was 52.4 dB relative to measured detection thresholds at 1 kHz (i.e., 52.4 dB SL). Presentation level varied
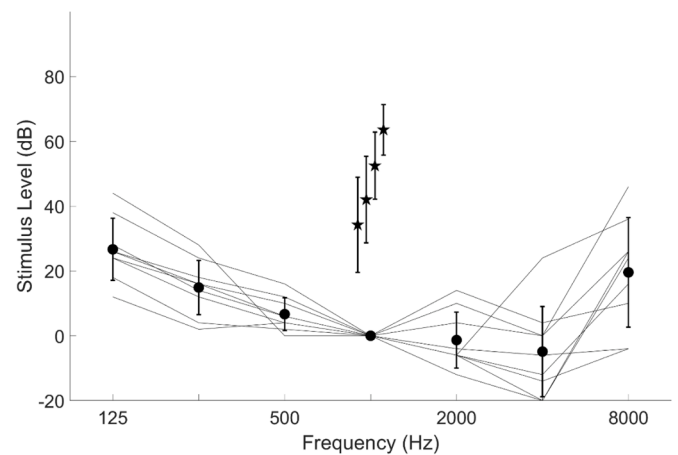

Fig. 3. Loudness scaling and detection thresholds for baseline calibration of presentation levels. All stimulus levels are normalized by detection thresholds for a 1 kHz pure tone. Detection thresholds for each subject are plotted as black lines with mean and standard deviations plotted as solid circles with error bars. Black stars with error bars indicate the average relative level and standard deviations that subjects set for a 1 kHz tone corresponding to "Soft", Medium Soft", "Medium", and "Medium Loud".

across subjects with a minimum of 40 and a maximum of 68 dB SL.

Consonant identification

Fig. 4 shows consonant identification scores as a function of SNR for the original and enhanced materials. In general, identification accuracy was better for the enhanced consonants compared to the original materials ($F_{1,12} = 241$, $p < 0.001$). The effect of SNR (as expected) was significant ($F_{2,12} = 230$, $p < 0.001$). The interaction between enhancement and SNR was also significant ($F_{2,12} = 232$, $p < 0.001$). Planned multiple comparisons indicated that pair-wise comparisons between original and enhanced identification scores were significant for the +6, 0, and -6 dB SNR conditions ($p < 0.01$). Cohen's d was calculated for each pair-wise comparison as an indication of effect size. The effect sizes were 0.39, 0.69, and 0.85 as measured at +6, 0, and -6 dB SNR, respectively.

Consonant identification was analyzed to consider individual benefits. Fig. 5 shows the enhanced benefit defined as the improvement in consonant identification averaged across phonemes for each subject.
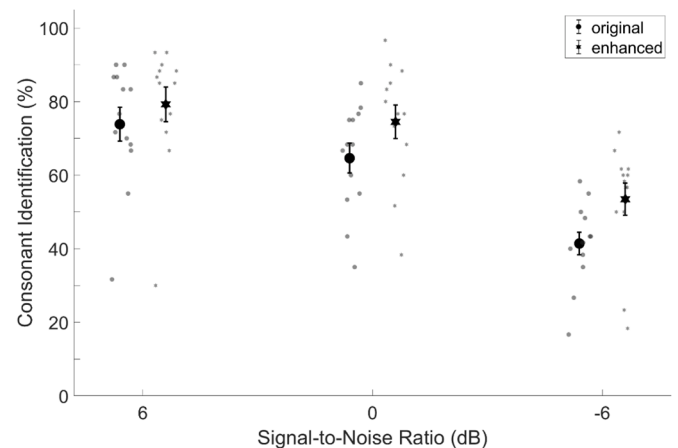

Fig. 4. Consonant identification as a function of SNR for the original and enhanced materials. Smaller gray symbols indicate identification accuracy of each subject. Larger black symbols indicate across subject averages and error bars indicate standard errors.
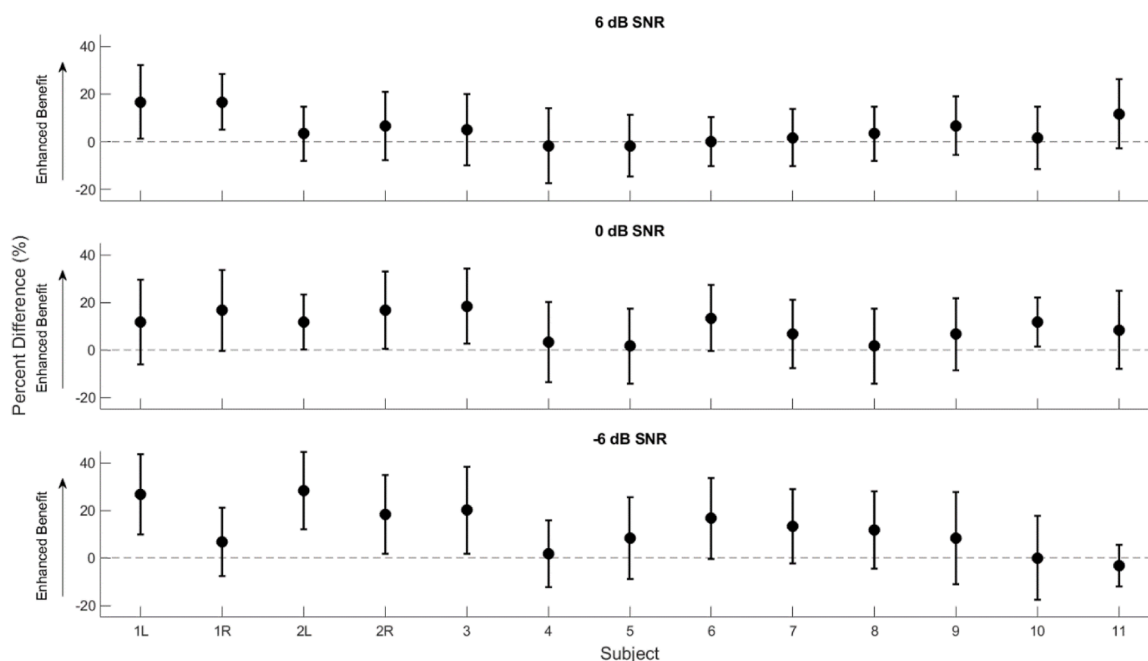
Fig. 5. Effect of audibility emphasis on consonant identification for each subject and SNR averaged across consonants. Error bars indicate standard deviations. The dotted line indicates no benefit, positive values indicate better performance with the enhanced consonants compared to the original.

While there was individual variability, almost all subjects for each SNR performed better with the enhanced materials, though this benefit rarely exceeded the standard deviation of this small sample.

Consonant identification was also analyzed to consider enhancement effects on a consonant-by-consonant basis averaged across subjects. Fig. 6 shows the enhanced benefit defined as the improvement in consonant identification averaged across subjects comparing enhanced versus original consonant materials. With decreasing SNR, the clear trend is that an enhancement benefit emerges for every consonant with a few notable exceptions. For the 0 and 6 dB SNR conditions, identification of "ACHA" was degraded by the emphasis algorithm. In contrast, the largest consistent improvements in consonant identification were observed for "ABA", "AGA", "ALA", "ARA", and "AWA". This noted pattern of degradation for "ACHA" but largest benefit for voiced stops and sonorants motivates future consideration for tailoring the algorithm based on specific consonant acoustics, noting the input intensity and input voicing strength.
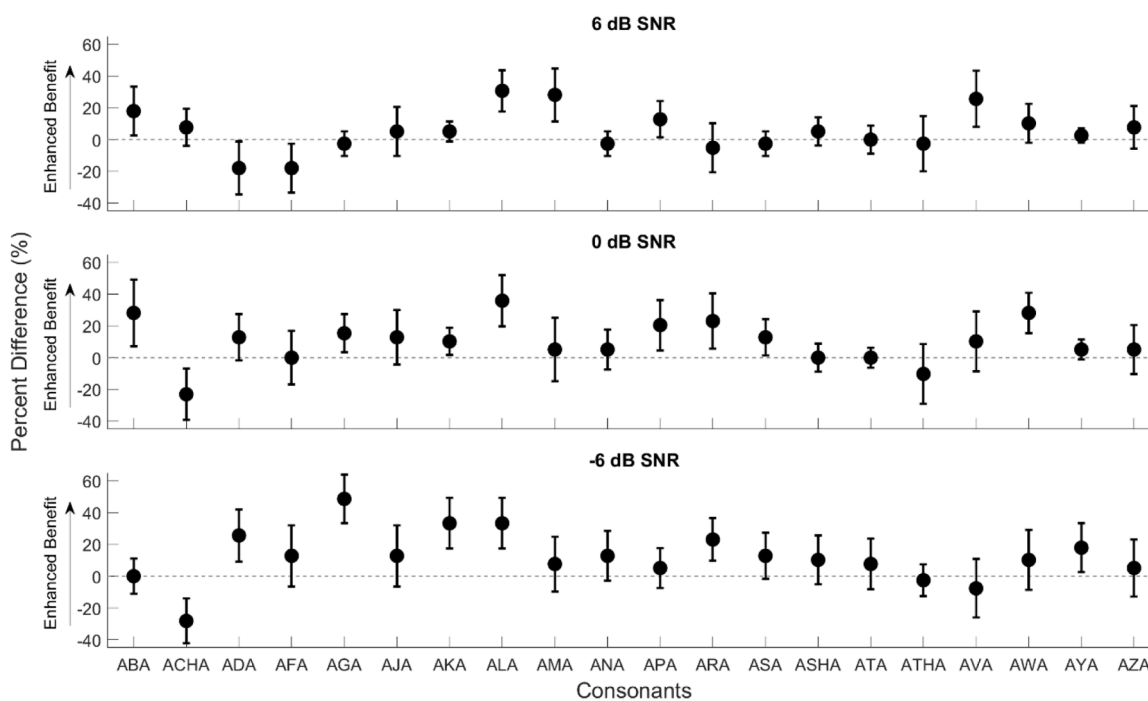

Fig. 6. Effect of audibility emphasis on consonant identification for each consonant. Solid circles indicate the difference in consonant identification between the enhanced and original materials averaged across subjects. Error bars indicate standard errors. The dotted line indicates no benefit, positive values indicate better performance with the enhanced materials compared to the original, and negative values indicate poorer performance with the enhanced materials.

Vowel identification

Fig. 7 shows vowel identification scores as a function of SNR for the original and enhanced materials. The effect of processing was not significant ($F_{1,12} = 0.06$, $p = 0.82$). While the effect of SNR was significant ($F_{2,12} = 88.9$, $p < 0.001$), the interaction between processing and SNR was not ($F_{2,12} = 0.18$, $p = 0.89$). Planned multiple comparisons indicated that no pair-wise comparison between original and enhanced identification scores was significant at any SNR tested ($p > 0.05$).

Vowel identification was analyzed to consider enhancement effects on a vowel-by-vowel basis like the analysis done for consonant identification. While there was no average effect of the enhancement algorithm on vowel identification, individual vowel contrasts were analyzed to determine if specific vowels were affected. The analysis indicated that no vowel-by-vowel comparison indicated a significant effect of processing condition ($p > 0.05$).

In summary, the evidence suggests that applying this audibility emphasis algorithm as a pre-processing strategy to utterances improves consonant identification without negatively affecting vowel identification.

DISCUSSION

The study described in this article considered a preprocessing algorithm for consonant enhancement. The experimental design focused on the situation where the target speech was available for processing separately from the background noise as would occur in telecommunications. The results clearly show that emphasizing low-level sounds improves consonant identification for cochlear implant users while not significantly affecting vowel identification. This finding is particularly important since consonant identification typically deteriorates more rapidly in cochlear implant users compared to vowel identification (Goldsworthy et al., 2013; Goldsworthy 2015). The immediate relevance of this experiment is that it suggests applying such preprocessing to telecommunications would improve speech comprehension for cochlear implant users, thus improving their quality of life.

The extended relevance of these findings is that audibility emphasis might be used to improve cochlear implant outcomes more broadly. In normal hearing, it has been argued that there is an evolutionary benefit from enhanced loudness of initial transient sounds, and that this benefit could be increased by the existence of both peripheral and central mechanisms (Rhode and Smith 1986; Golding et al., 1995). Studies of neural response in cochlear implant users suggest that the time-constant of effective temporal loudness summation is shorter compared to that

observed in normal hearing, which should influence the design of speech processing strategies (Cohen 2009). To replicate normal temporal summation behavior with cochlear implant stimulation, it may be necessary to progressively attenuate stimulation in a manner that emulates neural adaptation (Eggermont 2015; He et al., 2016).

The few studies of audibility and/or transient emphasis that have been conducted with cochlear implant users have reported improvements in speech perception. The transient emphasis spectral maximum approach (TESM: Vandali, 2001) applies gain to spectro-temporal components exhibited rapid changes (~20 ms). This approach has been shown to improve speech reception in multiple-talker babble noise and in combination with spectral expansion (Bhattacharya et al., 2011). However, no benefits were found when the input speech was at a relatively high 75 dB sound pressure level (Holden et al., 2005). These results indicate that the benefit from transient emphasis is thus connected to basic audibility of the low-level consonant cues.

On a more granular level, others have pointed out that the relatively low input level of consonants may negatively affect how consonants are processed by existing strategies for cochlear implants (Koning and Wouters 2016). Specifically, most existing sound processing strategies for cochlear implants select a subset of electrodes to stimulate during short time segments based on the input energy in the corresponding frequency regions. Thus, for low-level sounds, the issue of sensory encoding can be acerbated in that spectro-temporal components may be completely removed by electrode selection routines ("N-of-M selection") that only encode the strongest spectral regions. Emphasis of consonant acoustics prior to processing increases the likelihood that low-level consonants will be selected and encoded by such electrode selection routines. As argued by Koning and Wouters, 2016, this occurs because amplification of onsets leads to a higher probability that channels containing onset cues are selected. They further note that this interaction between consonant emphasis and channel selection would be most pronounced in the presence of background noise since maxima selection would more likely pick channels that contain speech information in comparison to channels dominated by noise.

More generally, a distinction should be made between audibility emphasis as presented here and adaptive dynamic range optimization algorithms that seek to encode low-level sounds. Adaptive dynamic range optimization algorithms have been developed that continually monitor input acoustics and amplify low-level sounds in the absence of any relatively high-level sounds in the environment (James et al., 2002; Blamey 2005). These algorithms employ spectral analysis and scan the outputs of spectral channels to determine if sound is being represented in each channel over a relatively long-term (seconds compared to the millisecond analyses used for transient and onset enhancement algorithms). Adaptive dynamic range optimization thus works to continually adjust the gain of spectral regions to ensure that input acoustics are encoded into audible stimulation for each region over a relatively long averaging time. This sort of processing has been shown to be effective for adjusting frequency region gain values when changing listening environments or compensating for varying input levels of speech; however, such adaptive dynamic range processing does not specifically enhance transient or onset acoustics. Hypothetically, then, the two styles of processing could be implemented together to both set the long-term statistical average of the frequency-specific gain values while also providing short-term gain enhancement for transient sounds as needed.

The results of the present study indicate that combining emphasis of low-level sounds with the type of fuzzy logic used with adaptive dynamic range optimization algorithms would likely yield synergistic benefits. The present study found that audibility emphasis was not needed in quiet when phonemes were presented at a comfortable listening level, but the benefit emerged when testing at progressively more difficult signal-to-noise ratios. Previous studies have found that audibility and/or transient emphasis can improve performance in quiet for low-level speech, but that the benefit diminishes once speech is at a sufficiently high input level (Firszt et al., 2004; Bhattacharya and Zeng
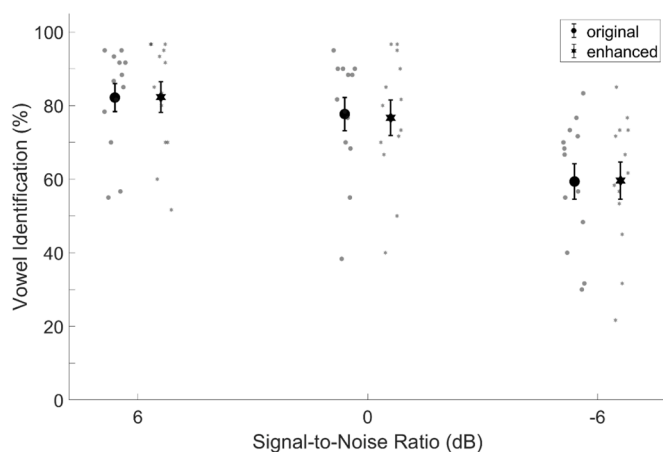

Fig. 7. Vowel identification as a function of SNR for the original and enhanced materials. Smaller gray symbols indicate identification accuracy of each subject. Larger black symbols indicate across subject averages and error bars indicate standard errors.

2007; Vandali and Clark 2010). A finding of the present study was that transient emphasis, while providing a large benefit for most consonants, particularly for sonorants, degraded identification of the consonant "ch" (/tʃ/). This finding may indicate that relatively loud consonants do not benefit from additional gain. Taken together, this suggests that transient emphasis could be optimized by considering the long-term input level, the input signal-to-noise ratio, as well as considering the input level of the sound, thus providing emphasis in a nuanced manner.

A limitation of the present study was that audibility enhancement of low-level sounds was not directly compared to simply increasing the overall gain. The advantage of only increasing the gain of low-level sounds is that it allows the listener to listen to the target speech at a comfortable level. The benefits observed in this study indicate that consonant identification can be improved at a specific SNR without requiring the listener to increase the overall gain. Nevertheless, it would be informative to directly compare phoneme identification for the enhancement strategy described here with an overall boost in gain for the target. Doing so would clarify the contributions of audibility, SNR, and overall presentation level, which would better inform algorithm optimization.

Another important consideration for applications of audibility emphasis for hearing devices is that most hearing devices support wireless streaming of audio from media devices to the hearing device. Such capacity allows a clean target signal to be transmitted with clarity while suppressing the ambient environment as needed. This capacity, in the extreme, allows the listener to mute the ambient environment if desired. Doing so circumvents the need for target-speech enhancement since the ambient noise is completely suppressed. However, muting the ambient environment comes with the downside that the listener's own voice will also be muted, which can be disconcerting. Preprocessing approaches for telecommunications will likely need to combine voice detection from the ambient environment with audibility emphasis on the incoming sound to optimally combine target enhancement with ambient noise suppression.

In conclusion, the results of the present study show that audibility emphasis of low-level sounds improves consonant identification while not significantly affecting vowel identification when the target speech is separately processed. These findings are relevant to telecommunications (e.g., media listening and watching, phone calls, video conferencing) since the transmitted audio is available on the listener's media device and can be enhanced prior to playback or audio streaming to the listener's hearing device. Such telecommunication applications have been consistently important to people with hearing loss but were made more so during the response to the COVID-19 pandemic. The results indicate that the benefits provided by audibility emphasis are large for the more adverse signal-to-noise ratios tested (0 and -6 dB). Providing audibility emphasis for telecommunications likely will lead to better speech comprehension for cochlear implant users, thus facilitating ease of communication and improving quality of life.

AUTHORSHIP STATEMENT

Audibility emphasis of low-level sounds improves consonant identification for cochlear implant users

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in Speech Communication.

Declaration of Competing Interest

The authors declares that they have no conflicts of interest.

REFERENCES

Baker, MW, Sarpeshkar, R, 2006. Low-Power Single-Loop and Dual-Loop AGCs for Bionic Ears. IEEE J Solid-State Circuits 41, 1983–1996.

Bhattacharya, A, Vandali, A, Zeng, F-G, 2011. Combined spectral and temporal enhancement to improve cochlear-implant speech perception. J Acoust Soc Am 130, 2951–2960. https://doi.org/10.1121/1.3641401.

Bhattacharya, A, Zeng, F-G, 2007. Companding to improve cochlear-implant speech recognition in speech-shaped noise. J Acoust Soc Am 122, 1079–1089. https://doi.org/10.1121/1.2749710.

Blamey, PJ, 2005. Adaptive Dynamic Range Optimization (ADRO): A Digital Amplification Strategy for Hearing Aids and Cochlear Implants. Trends Amplif 9, 77–98. https://doi.org/10.1177/108471380500900203.

Choi, SJ, Kim, JH, Kim, KH, 2008. Comparison of Speech Onset Detection Characteristics of Adaptation Algorithms for Cochlear Implant Speech Processor. J Biomed Eng Res 29, 25–31.

Cohen, J, 1992. A Power Primer. Quant Methods Psychol 112, 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Cohen, LT, 2009. Practical model description of peripheral neural excitation in cochlear implant recipients: 3. ECAP during bursts and loudness as function of burst duration. Hear Res 247, 112–121. https://doi.org/10.1016/j.heares.2008.11.002.

Delgutte, B, Kiang, NYS, 1984. Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. J Acoust Soc Am 75, 897–907. https://doi.org/10.1121/1.390599.

Desloge, JG, Reed, CM, Braida, LD, et al., 2017. Masking release for hearing-impaired listeners: The effect of increased audibility through reduction of amplitude variability. J Acoust Soc Am 141, 4452–4465. https://doi.org/10.1121/1.4985186.

Dreschler, WA, Plomp, R, 1985. Relations between psychophysical data and speech perception for hearing-impaired subjects. II. J Acoust Soc Am 78, 1261–1270.

Drullman, R, Festen, JM, Plomp, R, 1994. Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95, 2670–2680.

Eggermont, JJ, 2015. Animal models of auditory temporal processing. Int J Psychophysiol 95, 202–215. https://doi.org/10.1016/j.ijpsycho.2014.03.011.

Firszt, JB, Holden, LK, Skinner, MW, et al., 2004. Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. Ear Hear 25, 375–387. https://doi.org/10.1097/01.AUD.0000134552.22205.EE.

Friesen, LM, Shannon, R.V., Başkent, D, Wang, X, 2001. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J Acoust Soc Am 110, 1150. https://doi.org/10.1121/1.1381538.

Golding, NL, Robertson, D, Oertel, D, 1995. Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision. J Neurosci 15, 3138–3153. https://doi.org/10.1523/jneurosci.15-04-03138.1995.

Goldsworthy, RL, 2015. Correlations Between Pitch and Phoneme Perception in Cochlear Implant Users and Their Normal Hearing Peers. J Assoc Res Otolaryngol. https://doi.org/10.1007/s10162-015-0541-9.

Goldsworthy, RL, Delhorne, L.a., Braida, LD, Reed, CM, 2013. Psychoacoustic and Phoneme Identification Measures in Cochlear-Implant and Normal-Hearing Listeners. Trends Amplif 17, 27–44. https://doi.org/10.1177/1084713813477244.

He S, Abbas PJ, Doyle DV, et al (2016) Temporal Response Properties of the Auditory Nerve in Implanted Children with Auditory Neuropathy Spectrum Disorder and Implanted Children with Sensorineural Hearing Loss.

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

Hochberg, I, a, Boothroyd, Weiss, M, Hellman, S, 1992. Effects of noise and noise suppression on speech perception by cochlear implant users. Ear Hear 13, 263–271.

Holden, LK, Vandali, AE, Skinner, MW, et al., 2005. Speech recognition with the advanced combination encoder and transient emphasis spectral maxima strategies in nucleus 24 recipients. J Speech, Lang Hear Res 48, 681–701. https://doi.org/10.1044/1092-4388(2005/047.

James, CJ, Blamey, PJ, Martin, L, et al., 2002. Adaptive dynamic range optimization for cochlear implants: a preliminary study. Ear Hear 23, 49S–58S.

Koning, R, Wouters, J, 2016. Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users. Hear Res 342, 13–22. https://doi.org/10.1016/j.heares.2016.09.002.

Koning, R, Wouters, J, 2012. The potential of onset enhancement for increased speech intelligibility in auditory prostheses. J Acoust Soc Am 132, 2569–2581. https://doi.org/10.1121/1.4748965.

Niparko, JK, Tobey, EA, Thal, DJ, et al., 2010. Spoken language development in children following cochlear implantation. JAMA - J Am Med Assoc 303, 1498–1506. https://doi.org/10.1001/jama.2010.451.

Oppenheim, A.V., 1978. Applications of digital signal processing. Prentice Hall, Englewood Cliffs, NJ.

Rhode, WS, Smith, PH, 1986. Encoding timing and intensity in the ventral cochlear nucleus of the cat. J Neurophysiol 56, 261–286. https://doi.org/10.1152/jn.1986.56.2.261.

Shannon, R.V., Jensvold, A, Padilla, M, et al., 1999. Consonant recordings for speech testing. J Acoust Soc Am 106, L71–L74.

Smith, RL, 1979. Adaptation, saturation, and physiological masking in single auditory-nerve fibers. J Acoust Soc Am 65, 166–178. https://doi.org/10.1121/1.382260.

Steeneken, HJM, Houtgast, T, 2002. Phoneme-group specific octave-band weights in predicting speech intelligibility. Speech Commun 38, 399–411. https://doi.org/10.1016/S0167-6393(02)00011-0.

Swaminathan, J, Balachandran, R, Musacchia, G, 2020. Effects of speech enhancement on brainstem coding of consonants in normal-hearing listeners. In: Association for Research in Otolaryngology - Midwinter Meeting. San Jose, California.

Vandali, A, Clark, GM, 2010. Emphasis of short-duration transient speech features. J Acoust Soc Am 128, 2259.

Wilson, BS, Dorman, MF, 2008. Cochlear implants: A remarkable past and a brilliant future. Hear Res 242, 3–21. https://doi.org/10.1016/j.heares.2008.06.005.

Won, JH, Drennan, WR, Nie, K, et al., 2011. Acoustic temporal modulation detection and speech perception in cochlear implant listeners. J Acoust Soc Am 376. https://doi.org/10.1121/1.3592521.

Zeng, F-G, Grant, G, Niparko, JK, et al., 2002. Speech dynamic range and its effect on cochlear implant performance. J Acoust Soc Am 111, 377–386. https://doi.org/10.1121/1.1423926.

Zeng, F-G, Kong, Y-Y, Michalewski, HJ, Starr, A, 2005. Perceptual consequences of disrupted auditory nerve activity. J Neurophysiol 93, 3050–3063. https://doi.org/10.1152/jn.00985.2004.