



Speech Modification for Intelligibility in Cochlear Implant Listeners: Individual Effects of Vowel- and Consonant-Boosting

Juliana N. Saba¹, John H. L. Hansen¹

¹Center for Robust Speech Systems (CRSS), Cochlear Implant Processing Lab (CILab),
Erik Jonsson School of Engineering & Computer Science,
University of Texas at Dallas; Richardson, Texas, USA
juliana.saba@utdallas.edu, john.hansen@utdallas.edu

Abstract

Previous research has demonstrated techniques to improve automatic speech recognition and speech-in-noise intelligibility for normal hearing (NH) and cochlear implant (CI) listeners by synthesizing Lombard Effect (LE) speech. In this study, we emulate and evaluate segment-specific modifications based on speech production characteristics observed in natural LE speech in order to improve intelligibility for CI listeners. Two speech processing approaches were designed to modify representation of vowels, consonants, and the combination using amplitude-based compression techniques in the “electric domain” – referring to the stimulation sequence delivered to the intracochlear electrode array that corresponds to the acoustic signal. Performance with CI listeners resulted in no significant difference using consonant-boosting and consonant- and vowel-boosting strategies with better representation of mid-frequency and high-frequency content corresponding to both formant and consonant structure, respectively. Spectral smearing and decreased amplitude variation were also observed which may have negatively impacted intelligibility. Segmental perturbations using a weighted logarithmic and sigmoid compression functions in this study demonstrated the ability to improve representation of frequency content but disrupted amplitude-based cues, regardless of comparable speech intelligibility. While there are an infinite number of acoustic domain modifications characterizing LE speech, this study demonstrates a basic framework for emulating segmental differences in the electric domain.

Index Terms: cochlear implants, Lombard Effect, formants, signal processing, compression, enhancement

1. Introduction

The natural modification in human speech production to compensate for a noisy listening environment is referred to as Lombard Effect (LE) [1]. In the field of automatic speech recognition (ASR), stressed speech, including LE speech, has been shown to degrade performance of ASR systems and have motivated both compensation as well as the need for additional training paradigms [2], [3]. In the field of speech science, investigations of speech understanding from natural LE speech have demonstrated positive benefits for speech-in-noise listening situations, both in normal hearing (NH) and hearing impaired listeners [4]–[9]. Various approaches for artificial perturbation have also demonstrated improvements in LE classification and have contributed to partial restoration of intelligibility deficits in challenging listening environments

[3], [6], [8], [10], [11]. Specific to the hearing impaired, cochlear implant (CI) listeners are able to achieve high speech intelligibility performance in quiet, noise-free conditions, however, the additive effects of noise, reverberation, and competing speakers/environments reduces effective communication [12]–[14]. In this study, we aim to leverage amplitude-based differences in LE speech to determine the effects of boosting consonants and vowels within the electric domain. Similar to the acoustic- or frequency-domain, the electric domain corresponds to the electrode stimulation patterns that define place and amplitude cues in normal hearing but with electrode and current information.

LE speech, or noise-induced speech, can be characterized by a number of speech production parameters such as a rise in pitch (F0), a more flattened glottal spectral slope, a rise of intensity for vowels and consonants, a corresponding increase in first (F1) and second (F2) formants, and segmental shifts in duration [3]–[5], [15]–[18]. To consider artificially reproducing LE speech, Cairns and Hansen identified the various domains of LE-like modifications and the amount of modifications for each domain in a framework known as Source Generator theory [19]. While explicit testing of the infinite number of combinational perturbation generators is cumbersome, artificial perturbation using individual or subsets of these perturbation domains have demonstrated benefits in intelligibility and ASR performance [10], [11], [20].

Integration of pre-emphasis, speech enhancement, and noise suppression techniques within the CI signal processing pipeline can improve intelligibility. However, these approaches focus on the noisy signal or removal of problematic features instead of boosting the target signal above the noise floor. It is possible to evaluate the effects of individual and combinational modification approaches proposed to generate synthetic or artificial LE speech. Previously, studies of intelligibility benefits from F0-modified speech, speech with adjusted spectral tilt, temporal-based, amplitude-modified speech, and duration-modified speech have provided insights on what acoustic features can drive the input signal towards a more Lombard-like output [4]–[6], [17], [21]–[23]. Unlike previous studies focusing on speech modification using a combinational approach, here we evaluate an amplitude-based, single modification approach to improve speech intelligibility for CI listeners. Custom compression functions are designed to improve the electric representation of speech by targeting individual segments, (i.e., consonant and vowel regions). We hypothesize that modifications evaluated individually will highlight how successful LE perturbation for a single domain can emulate benefits observed in multi-domain approaches.

2. Signal Processing Approaches

A baseline ‘n’-of-‘m’ strategy, the Advanced Combination Encoding (ACE) from Cochlear Ltd., is used as the pipeline for each of the proposed CI signal processing approaches [24]. In this commercial strategy, the signal is pre-emphasized and decomposed into time and frequency components using a 22-channel filter bank associated with a 22-electrode implant system. The envelope of each filter band is extracted and energy is used to determine ‘n-maxima’ or the ‘n’ channels with the highest energy. ‘n’ filter outputs are then passed through a logarithmic compression function, known as the Logarithmic Growth Function (LGF) defined as,

$$\text{LGF} = \frac{\ln\left(1 + \alpha_{\text{LGF}} \frac{(x_{\text{PSD}} - \text{BL})}{(\text{SL} - \text{BL})}\right)}{\ln(1 + \alpha_{\text{LGF}})} \quad (1)$$

where power spectral densities, x_{PSD} , are mapped to clinical units represented as a percentage of the dynamic range (DR) of each electrode and α_{LGF} is based on the loudness growth function. Values of x_{PSD} are normalized according to base (BL) and saturation (SL) limits of intracochlear electrodes, respectively. Post-LGF values greater than SL are set to 1 and values below BL are set to 0. The DR is constrained between threshold (THR) and maximum comfort levels (MCL) according to a clinical MAP (configuration of electric parameters individualized to hearing loss in CI recipients set by audiologists). Lastly, biphasic pulses are constructed using current (amplitude) and electrode (frequency) information and transmitted to an internal radio frequency receiver which delivers the electric signal to the intracochlear electrode array.

Previously, a channel selection criteria was used to prioritize ‘l’ channels associated with formant frequencies in order to decrease selection of noise-dominant bands [25]. Using the same criteria, various compression functions and selection criterion are integrated to form two proposed CI signal processing strategies which individually boost consonants, and vowels and consonants, in addition to prioritization and boosting of formants demonstrated in [25].

Spectral structure of consonants in the presence of noise, unfortunately, is not conducive to a selection strategy based on energy. To overcome this, two components are integrated within the ‘n’-of-‘m’ framework to enhance the electric representation of consonants in a strategy referred to as the Consonant Boosting Strategy (CBS): (i) customized frequency tables, and (ii) sigmoid compression function. To modify the representation of vowels, a sharp compression function is used to increase the input/output relationship of vowels in addition to changes made to consonant segments in a strategy referred to as Consonant and Vowel boosting via Sigmoid and Weighted Logarithmic compression (CVSWL).

2.1. Consonant-boosting Approach

In the CBS speech processing approach, channel selection criterion is influenced by the generalized electric representation or time-frequency responses specific to each consonant referred to as consonant frequency tables. Like the spectrogram, an electrodiagram represents “electric hearing” where electrodes are used to represent frequency and singular vertical lines are used to represent current instead of amplitudes demonstrated with differing contrast in the acoustic domain. Electrodes are inversely proportional to channels such that low-frequency channels are represented as higher electrodes (towards the apex of the cochlea) and high-

frequency channels are represented as lower electrodes (towards the base of the cochlea). Offline phonetic transcription is used to identify voiced and unvoiced frames to select channels accordingly. For voiced frames, the channel selection process uses the ‘l’-of-‘n’-of-‘m’ criteria in [25] to select and boost formants where the remaining ‘n-l’ channels are sent to the LGF in Eq. (1). For unvoiced frames, channels with the highest spectral energy are compared against consonant frequency tables which define the spectral representation of 19 consonants observed in quiet, noise-free speech in the electric domain. If more than half of the channels selected using the ‘n-maxima’ criteria are different than channels identified for the phoneme, the selection is updated to include channels defined in consonant frequency table. After selection, these channels are subjected to the sigmoid gain function (SGF) described in Eq. (2).

2.1.1. Customized Consonant Frequency Tables

Frequency tables were generated by analyzing the following consonants in the /VCV/ (vowel-consonant-vowel) database: /b/, /ch/, /d/, /dh/, /g/, /jh/, /k/, /l/, /m/, /n/, /p/, /r/, /s/, /sh/, /t/, /v/, /w/, /y/, /z/ [26]. Consonant tokens were processed using the same 22-channel filter bank with default frequency allocations set by Cochlear Ltd. averaged across male and female speakers. x_{PSD} values were used to identify spectral patterns based on amplitude, tonotopic organization, and the number of distinct and consecutive or adjacent channels.

Stimulation/selection patterns were generalized using the ‘n-maxima’ approach to determine selection for each consonant for commonly used ‘n’ values observed in clinical MAPs, (i.e., 8, 10, and 12). Selection was characterized by the number of clustered (adjacent) channels (single, double, triple) and by center frequencies of each filter band (high, low, mid). The following types of patterns were generalized: (a) high, (b) high-double, (c) high-single, (d) high-triple, (e) hybrid, (f) low, (g) low-double, (h) low-single, and (i) mid. For example, “ASA” has a distinct high frequency region of energy for channels 18–22 (~4–8 kHz), so the identified pattern is ‘high’.

2.1.2. Sigmoid Gain Function (SGF)

A sigmoid relationship was designed to boost low-level energies of unvoiced consonants. In an offline analysis, average current and electrode behavior from phonetically labeled consonants of 450 IEEE sentences were used to determine electric representation of each phoneme. More than 2/3rds of x_{PSD} values did not meet BL criteria for stimulation such that frequency information was not provided to the implant listener using the traditional LGF compression function in ACE. Therefore, α_{SGF} and β_{SGF} values in the sigmoid function were defined where minimum stimulation represents 33% DR and midpoint of the linear segment for the sigmoid function represents the median value of x_{PSD} observed in the consonant analysis. The SGF relationship is defined as,

$$\text{SGF} = \frac{1}{1 + e^{(\alpha_{\text{SGF}} x_{\text{PSD}} + \beta_{\text{SGF}})}} \quad (2)$$

where scaling factors for α_{SGF} and β_{SGF} are 29.0415 and -1.1015, respectively. All values below SL follow the linear portion of the sigmoid function to stimulate low-level segments of unvoiced features of speech.

2.2. Vowel- and Consonant-boosting Approach

In the CVSWL speech processing approach, channel selection is individualized according to offline phonetic transcription for

voiced and unvoiced frames. For unvoiced frames, the selection and compression functions defined in the CBS strategy are used. For the voiced frames, formant channels are prioritized first prior to ‘n-maxima’ selection as in [25]. Additional boosting is applied in an unconstrained, logarithmic manner where energy is not preserved within the frame using a weighted LGF defined in Eq. (3).

2.2.1. Weighted Logarithmic Compression Function (LGF_w)

To perform vowel boosting, the output of ‘n-maxima’ filter bands are processed through a weighted LGF function (LGF_w) described as,

$$LGF_w = \sqrt{\frac{\ln\left(1 + \alpha_{LGF} \frac{(\alpha_{PSD-BL})}{(SL-BL)}\right)}{\ln(1 + \alpha_{LGF})}} \quad (3)$$

where the LGF function from Eq. (1) is raised to the power of 0.5 (determined in an offline analysis as the optimal weighting factor) such that the midpoint of the SGF function represents 50% of the DR to increase amplitude cues for vowels.

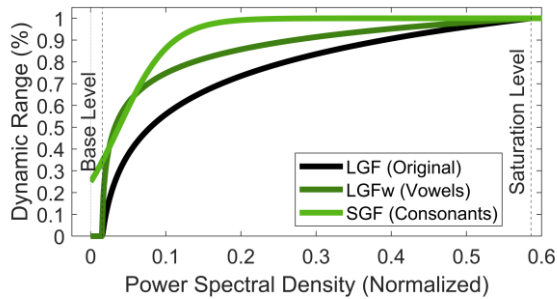


Figure 1. Proposed compression functions for vowels (dark green) and consonants (light green).

3. Evaluation and Results

3.1. Experimental Protocol

Speech intelligibility was evaluated in an acute listening experiment with 4 CI users in quiet and two simulated noisy conditions using additive speech-shaped-noise (SSN). Sentences from the IEEE database were used to determine the number of words correctly identified in quiet, +10, and +5 dB signal-to-noise ratio (SNR). Stimuli was randomized across condition and strategy and provided to CI listeners in a direct-connect manner (i.e., by-passing clinical processor through radio frequency transmission of electric stimuli) using the UT-Dallas CCI-MOBILE Research Platform [27], [28].

The experimental protocol was completed in two phases using two different groups of CI users with Cochlear Ltd. implant systems using ACE as the clinical signal processing strategy. In the first phase, the control (ACE) and consonant-boosting (CBS) approaches were evaluated with 2 CI subjects. The bilateral subject was asked to test each ear independently in a unilateral mode, (i.e., N=3). In the second phase, the control (ACE) was evaluated with 2 different CI subjects against the consonant and vowel-boosting strategy (CVSWL).

3.2. Intelligibility Results with CBS (Consonant-boosting)

Speech intelligibility with CBS is shown in Fig. 2A. A two-way Analysis of Variance (ANOVA) revealed a significant effect of noise ($F[2,4]=117.4, p=0.0003$), but not for strategy

or interaction. Comparable performance was observed between ACE and CBS ($p>0.05$), where CBS was 8.2%, 9.3%, and 3.3% points below ACE for quiet, +10, and +5 dB SNR SSN, respectively. Individual improvements of +3.2% points for S2-L at +10 dB SNR and +7.78% points for S1 at +5 dB SNR were observed. S1 and S2 evaluated with CBS are considered low performing according to an average baseline of 47.6% with their clinical processing strategy in quiet.

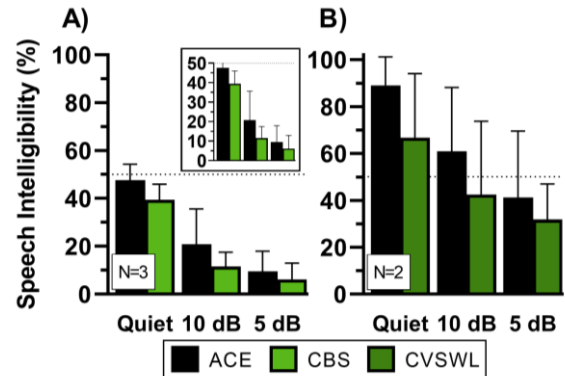


Figure 2. Speech intelligibility (%) of CI listeners for A) CBS (N=3), and B) CVSWL (N=2) compared to the baseline (ACE) strategy, where error bars indicate standard deviation.

3.3. Intelligibility Results with CVSWL (Consonant and Vowel Boosting)

Fig. 2B illustrates average speech intelligibility (N=2) with the CVSWL strategy. Results from a 2-way ANOVA also revealed a significant effect of noise ($F[2,2]=34.48, p=0.0282$), but not for strategy or interaction. Unlike CBS, individual performance for S3 and S4 did not result in improvement across the three conditions tested. Average speech intelligibility with CVSWL was 22.3%, 18.5%, and 9.5% lower than ACE for quiet, +10, and +5 dB SNR SSN. Post-hoc comparisons tests failed to reach significance between the two strategies ($p>0.05$). Individual performance for S4 increased as SNR decreased, whereas the opposite effect occurred for the highest performing subject, S3. Both of these subjects (S3, S4) are considered high performing with average baseline performance of ACE at 89.0%. In both of the proposed strategies, as SNR decreased, the difference in performance between ACE also decreased.

3.4. Comparison of Electric Stimulation Patterns

Electrograms, representative of electrode and current stimulation across time, are shown in Fig. 3 for S3. The height of the vertical lines represents the amount of stimulated current across time and frequency, represented as electrodes. The proposed strategies delivered more apical stimulation (electrodes 16-22, 0.2-1 kHz) and contributed towards a fuller representation of voiced segments in the range of F2 (electrodes 8-12, 1.4-2.8 kHz). Additional stimulation of onset and offset cues can be seen with both the consonant and vowel boosting strategies (e.g., consonant-like segments: 400-500ms, 1000-1200ms, 1500-1600ms, and 1900-2100ms). Compared to ACE, the CVSWL strategy provided stimulation across the entire frequency spectrum in an interleaved manner instead to stimulating only a subset of the segment-based information.

Table 1 quantifies the electrical stimulation from the electrogram for S3 (Fig. 3). The slope represents a linear fit

of a histogram used to define the relationship between center frequencies (electrodes) and stimulated frames. Negative slopes indicate more stimulation of lower-frequencies, whereas positive slopes indicate more stimulation of higher-frequencies. While visual inspection of electrodograms suggest CBS and CVSWL strategies improved consonant representation, a 1.2-fold and 3.3-fold decrease in slope was observed. Average stimulated current and mathematical mode for both strategies was found to be higher than ACE. Interestingly, analysis of voiced frames indicated higher current with CBS than with CVSWL strategy, whereas the opposite effect was observed for unvoiced frames, where current contributed from vowel-boosting and consonant-boosting was more dominant, respectively.

Table 1. Stimulation statistics of the sentence, “Watch the log float in the wide river” using an electrodogram for S3.

Frames	Voiced			Unvoiced		
	ACE	CBS	CVSWL	ACE	CBS	CVSWL
Slope	-0.139	-0.198	-0.157	0.019	-0.005	-0.045
Min	123	125	127	120	125	127
Med.	159	169	165	142	154	161
Max	171	180	180	170	180	180
Mode	160	178	178	132	178	180
Mean	157.88	166.00	162.96	144.31	154.72	160.75

4. Discussion

Of the strategies aimed to emulate segmental amplitude changes identified in LE speech, more comparable performance to baseline was observed with consonant-boosting than combinational vowel- and consonant-boosting approach. This is consistent with previous literature where improving consonant structure was shown to improve intelligibility [5], [21], [29], [30]. The behavior of SGF saturates much more quickly than the logarithmic-based functions. Saturation in this manner may have contributed to higher amplitude-cues for low frequency electrodes quantified by increased average current for voiced and unvoiced frames. This phenomenon is illustrated by spectral smearing in electrodograms in Fig. 3. Increased electric representation of low-frequency information is prevalent in both proposed strategies which upholds negative spectral slope shown in Table 1 for CBS and CVSWL. Since more energy is present in lower frequencies, it is possible that the compression functions modified these cues such that boosting done in high frequency region was overshadowed. It should be noted that both proposed strategies improved the representation of formants using the prioritization strategy in [25].

Analysis of stimulated current revealed levels closest to MCL regardless of fuller representation of spoken words observed in the electrodogram. It is possible that boosting of adjacent electrodes at higher amplitudes defined in the constant frequency tables stimulates the same group of spiral ganglion cells, thus increasing the likelihood for channel interaction and refractory behavior known to degrade speech perception [31]–[34]. Acoustic modification in this manner may create a “boosted” percept where frequency information is smeared across time and frequency and is less Lombard-like where amplitude changes in time and frequency occur naturally [3], [4], [35]. Unlike previous perturbation methods [6], [8], [9], amplitude modifications did not explicitly use model or adaptation data from natural LE speech. Here, the aim was to determine the relationship between generalized, segmental amplitude-based modifications and intelligibility. CBS, the more successful of the proposed approaches, resulted in frequency-targeted stimulation of either predominantly high or predominantly low frequencies, whereas CVSWL resulted in stimulation across the full entire speech spectrum.

Only a single subset of mathematical components was used in the design of the proposed compression functions. There may exist other combinations of scaling factors yielding better performance and electric representation more closely resembling natural LE speech. This proof-of-concept study demonstrates the ability of compression and selection criterion within CI signal processing strategies to adapt stimulation patterns to mimic amplitude-based components in LE speech.

5. Conclusions

Two compression-based boosting approaches (embedded within the processing pipeline of a commercial CI signal processing strategy) were evaluated to determine the potential benefit of intelligibility improvements for CI listeners in noisy conditions. Comparable performance to the baseline was achieved, however, the proposed strategies improved electric representation of consonants. The lack of intelligibility benefits was attributed to loudness artifacts, spectral smearing, and disruption of amplitude-based cues. An increased representation of consonants and F1-F2 features was observed providing evidence to emulate a single domain of LE speech using acoustic-based compression techniques.

6. Acknowledgements

This work was supported by Grant No. R01-DC016839 from the National Institute of Health, National Institute on Deafness and Other Communication Disorders.

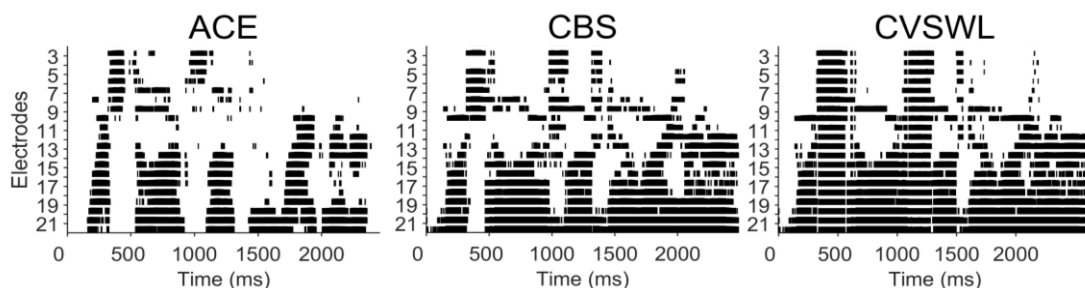


Figure 3. Electrodograms for the sentence, “Watch the log float in the wide river” with each signal processing strategy.

7. References

- [1] E. Lombard, "Le Signe de Lelevation de la Voix," *Annales des Maladies de l'Oreille, du Larynx, du Nez, et du Pharynx*, vol. 37, no. 2, pp. 101–119, 1911.
- [2] D. A. Cairns and H. L. Hansen, "Nonlinear analysis and classification of speech under stressed conditions," *The Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3392–3400, 1994.
- [3] J. H. L. Hansen, "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," *Speech Communication*, vol. 20, no. 1–2, pp. 151–173, 1996.
- [4] Y. Lu and M. Cooke, "Speech production modifications produced by competing talkers, babble, and stationary noise," *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 3261–3275, 2008.
- [5] Y. Lu and M. Cooke, "The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise," *Speech Communication*, vol. 51, no. 12, pp. 1253–1262, 2009.
- [6] J. H. L. Hansen, J. Lee, H. Ali, and J. N. Saba, "A speech perturbation strategy based on 'lombard effect' for enhanced intelligibility for cochlear implant listeners," *The Journal of the Acoustical Society of America*, vol. 147, no. 3, pp. 1418–1428, 2020.
- [7] M. Cooke, C. Mayo, and J. Villegas, "The contribution of durational and spectral changes to the Lombard speech intelligibility benefit," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 874–883, 2014.
- [8] J. N. Saba and J. H. L. Hansen, "The effects of Lombard perturbation on speech intelligibility in noise for normal hearing and cochlear implant listeners," *The Journal of the Acoustical Society of America*, vol. 151, no. 2, pp. 1007–1021, 2022.
- [9] J. N. Saba, "Leveraging Landmark Acoustic Features in Cochlear Implant Signal Processing," Ph.D. Dissertation, School of Engineering and Computer Science, University of Texas at Dallas, Richardson, TX, 2021.
- [10] S. E. Bou-Ghazale and J. H. L. Hansen, "Generating stressed speech from neutral speech using a modified CELP vocoder," *Speech Communication*, vol. 20, no. 1–2, pp. 93–110, 1996.
- [11] J. H. L. Hansen and S. E. Bou-Ghazale, "Robust Speech Recognition Training via Duration and Spectral-Based Stress Token Generation," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 5, pp. 415–421, 1995.
- [12] C. M. McKay, H. J. McDermott, A. E. Vandali, and G. M. Clark, "A comparison of speech perception of cochlear implantees using the spectral maxima sound processor (SMSP) and the MSP (MULTIPEAK) processor," *Acta Oto-Laryngologica*, vol. 112, no. 2, pp. 752–761, 1992.
- [13] M. W. Skinner, P. L. Arndt, and S. J. Staller, "Nucleus® 24 Advanced Encoder Conversion Study: Performance versus preference," *Ear and Hearing*, vol. 23, no. 1 SUPPL., pp. 2–17, 2002.
- [14] M. W. Skinner, L. K. Holden, L. A. Whitford, K. L. Plant, C. Psarros, and T. A. Holden, "Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," *Ear and Hearing*, vol. 23, no. 3, pp. 207–223, 2002.
- [15] J. C. Junqua, "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex," *Speech Communication*, vol. 20, no. 1–2, pp. 13–22, 1996.
- [16] J. C. Junqua, "The lombard reflex and its role on human listeners and automatic speech recognizers," *The Journal of the Acoustical Society of America*, vol. 93, no. 1, pp. 510–524, 1993.
- [17] M. Cooke, C. Mayo, C. Valentini-Botinhao, Y. Stylianou, B. Sauert, and Y. Tang, "Evaluating the intelligibility benefit of speech modifications in known noise conditions," *Speech Communication*, vol. 55, no. 4, pp. 572–585, 2013.
- [18] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *The Journal of the Acoustical Society of America*, vol. 84, no. 3, pp. 917–928, 1988.
- [19] J. H. L. Hansen and D. A. Cairns, "ICARUS: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments," *Speech Communication*, vol. 16, no. 4, pp. 391–422, 1995.
- [20] S. E. Bou-Ghazale and J. H. L. Hansen, "HMM-based stressed speech modeling with application to improved synthesis and recognition of isolated speech under stress," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 3, pp. 201–216, 1998.
- [21] M. Garnier and N. Henrich, "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?" *Computer Speech and Language*, vol. 28, no. 2, pp. 580–597, 2014.
- [22] E. Godoy, M. Koutsogiannaki, and Y. Stylianou, "Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles," *Computer Speech and Language*, vol. 28, no. 2, pp. 629–647, 2014.
- [23] E. Godoy, O. Rosec, and T. Chonavel, "Voice conversion using dynamic frequency warping with amplitude scaling, for parallel or nonparallel corpora," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 4, pp. 1313–1323, 2012.
- [24] A. E. Vandali, "Emphasis of short-duration acoustic speech cues for cochlear implant users," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2049–2061, 2001.
- [25] J. N. Saba, H. Ali, and J. H. L. Hansen, "Formant priority channel selection for an 'n-of m' sound processing strategy for cochlear implants," *The Journal of the Acoustical Society of America*, vol. 144, no. 6, pp. 3371–3380, 2018.
- [26] R. V. Shannon, A. Jensvold, M. Padilla, M. E. Robert, and X. Wang, "Consonant recordings for speech testing," *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. L71–L74, 1999.
- [27] R. Ghosh, H. Ali, and J. H. L. Hansen, "CCi-MOBILE: A Portable Real Time Speech Processing Platform for Cochlear Implant and Hearing Research," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1251–1263, 2021.
- [28] J. H. L. Hansen et al., "CCi-MOBILE: Design and Evaluation of a Cochlear Implant and Hearing Aid Research Platform for Speech Scientists and Engineers," 2019.
- [29] D. Fogerty and D. Kewley-Port, "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 847–857, 2009.
- [30] D. Kewley-Port, T. Z. Burkle, and J. H. Lee, "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2365–2375, 2007.
- [31] G. S. Stickney, P. C. Loizou, L. N. Mishra, P. F. Assmann, R. V. Shannon, and J. M. Opie, "Effects of electrode design and configuration on channel interactions," *Hearing Research*, vol. 211, no. 1–2, pp. 33–45, 2006.
- [32] C. M. McKay, M. D. Remine, and H. J. McDermott, "Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1514–1524, 2001.
- [33] S. Fredelake and V. Hohmann, "Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation," *Hearing Research*, vol. 287, no. 1–2, pp. 76–90, 2012.
- [34] Q. Fu, "Temporal processing and speech recognition in cochlear implant users," *Auditory and Vestibular Systems*, vol. 13, pp. 1635–1639, 2002.
- [35] J. Lee, H. Ali, A. Ziaei, E. A. Tobey, and J. H. L. Hansen, "The Lombard effect observed in speech produced by cochlear implant users in noisy environments: A naturalistic study," *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. 2788–2799, 2017.