I. INTRODUCTION

Cochlear implants (CI) provide the ability for hearing-impaired and/or deafened individuals to decode speech through electrical stimulation. Performance of CI users with commercial devices demonstrates speech recognition and speech understanding scores of 75% or higher in quiet conditions (Dorman et al., 1989; Dorman and Spahr, 2006; Dowell et al., 1986; Shannon et al., 1995; Skinner et al., 1991; Vandali et al., 2000). One of the many challenges in CI research is the decreased ability of users to achieve similar performance in the influence of noise. Due to the nature of CI-based processing, a reduction in temporal fine structure, low stimulation rates, decreased audio-visual cues, and increased amplitudes of noise across all channels may contribute to the decreased ability of hearing-impaired and CI users to decode speech (Assmann and Summerfield, 2004; Eddington, 1980; Fu et al., 1998; Kewley-Port et al., 2007; Lu and Cooke, 2008; Neuman et al., 2012; Summers et al., 1988; Zeng et al., 2005). During these unfavorable conditions, performance can be supplemented by providing the implant user with highly intelligible speech. Researchers have used speech enhancement and noise suppression techniques while others have employed speech modification techniques to improve CI user intelligibility.

Noise suppression strategies aim to remove noise from the input signal in order to identify the targeted segments of speech to yield higher intelligibility (Kokkinakis et al., 2012; Hu and Loizou, 2007, 2010; Ye et al., 2014). Speech enhancement schemes such as spectral subtraction, Wiener filtering, and other statistical techniques aim to construct an enhanced signal by deconstructing elements in the time versus frequency domains (Doclo et al., 2015; Hazrati et al., 2014; Loizou et al., 2005; Loizou, 2013; Yang and Fu, 2005). Speech modification techniques, however, are performed in the acoustic space to lift the target speech above the noise floor. Various approaches consist of increasing the signal-to-noise ratio (SNR), redistributing energy, adjusting or altering the time-domain characteristics, and adjusting speaking style (Donaldson and Allen, 2003; Kewley-Port et al., 2007; Lu and Cooke, 2008, 2009; Skinner et al., 1997). Past speech modification techniques have discovered additional approaches outside of the obvious solution of increasing the desired signal properties to generate greater separation from background noise, i.e., increasing the SNR. In CI commercial speech processors for implants manufactured by Cochlear Ltd., clinical unit limitations and other safe-guarding methods are employed to prevent the implant from over-stimulation and/or producing signals exceeding the CI user specific “maximum comfort level.”

All listeners, both normal and hearing-impaired, can naturally modify their speech in a response to noise exposure through a phenomenon referred to as the “Lombard effect” (LE) (Lombard, 1911). Acoustic components of Lombard speech differ from that of neutral speech. Lombard speech is generally associated with a more flattened spectrum with emphasis in high frequencies, longer
duration of target phonemes, slower speaking rate, and formant frequency adjustments (Bou-Ghazale and Hansen, 1996, 1997, 1998, 2000; Garnier et al., 2010; Godoy et al., 2014; Hansen 1988, 1989; Hansen and Bria, 1990; Junqua, 1992; Kewley-Port et al., 2007; Lee et al., 2015; Lee et al., 2017; Lu and Cooke, 2008). LE has been previously studied to assess its characteristics and further define the role of noise exposure on speech production and perception. A common approach to mimic LE speech is to redistribute energy across both time and frequency, trading off high frequency amplitude from low frequencies or redistributing the entire spectrum (Cooke et al., 2013; Cooke et al., 2014; Lu and Cooke, 2009; Niederjohn and Grotefueschen, 1976; Schepker et al., 2013; Zorila et al., 2012). Dynamic range compression (DRC) has also been used to enhance clean qualities of speech by reducing peaks, spectral tilt, and sharpen formant information (Zorila et al., 2012; Godoy and Stylianou, 2013). Durational aspects of speech have been investigated to determine its independent role on speech intelligibility (Lu and Cooke, 2009; Cooke et al., 2014). Optimization techniques based on the objective metrics such as the Speech Intelligibility Index (SII) have been used to adjust amplitude and compression ratios to produce desirable SNRs in comparison to neutral or unmodified speech (Godoy and Stylianou, 2012, 2013; Schepker et al., 2013). Overall, the role of spectral reorganization and other modification techniques inspired by LE speech have led to significant intelligibility benefits for normal hearing users, within ranges of 12%–42% percentage points in listener evaluations (Cooke et al., 2014; Godoy et al., 2014; Lu and Cooke, 2008, 2009; Schepker et al., 2013; Zorila et al., 2012) and increasing SII by 0.7–1.9 points (Godoy et al., 2014; Godoy and Stylianou, 2014). Other modifications of speaking style for applications such as text-to-speech, classification, or automatic speech recognition (ASR) systems aim to either synthesize speech or classify speaker stress/emotion (Boril and Hansen, 2010; Hansen, 1988, 1989, 1996; Hansen and Bria, 1990; Hansen and Varadarajan, 2009; Hansen and Womack, 1996; Hansen et al., 2011; Zhou and Hansen, 1998; Zhou et al., 1999, 2001). A speaker-independent perturbation model was integrated into an ASR model by Hansen and Cairns to define differences in duration, amplification, and spectral shape of Lombard speech to improve recognition in noise (Hansen and Cairns, 1995). Speech modifications have also been used to produce variations of stress or emotion such as angry, Lombard, neutral, loud, etc., through mathematical models in text-to-speech systems (Bou-Ghazale and Hansen, 1996, 1997, 1998, 2000; Hansen and Cairns, 1995). This instinctual modification and its potential benefits in speech intelligibility (SI) have been studied for human ASR and the general speech science community, however, little work has been done to investigate LE in the CI community.

In a previous study by the authors, it was shown for the first time that CI users were able to produce Lombard speech from a range of noise exposure types resulting in an increase in vocal effort, flattening of spectral slope, and increase in phoneme duration (Lee, 2017; Lee et al., 2015, 2017). Although hearing impaired, the ability of CI users to produce Lombard speech indicates the intact feedback system linking perception and production. The spectro-temporal changes defining Lombard speech have been identified and demonstrate intelligibility benefits for normal hearing individuals (Bou-Ghazale and Hansen, 1996, 1997, 1998, 2000; Lu and Cooke, 2008, 2009; Cooke et al., 2013). Additionally, Lombard-like modification approaches have also demonstrated significant intelligibility benefits, evaluated by speech systems and/or normal hearing users (Bou-Ghazale and Hansen, 1996, 1997, 1998, 2000; Cooke et al., 2013; Godoy and Stylianou, 2013; Godoy et al., 2014; Lu and Cooke, 2008, 2009; Zorila et al., 2012). Can CI users, whose speech perception is supplemented via electrical stimulation, perceive the same changes in speech production due to LE and leverage them to receive intelligibility benefits in the same way as normal hearing users? This study addresses the potential gains in speech intelligibility specifically for CI users from natural Lombard speech as well as artificially perturbed Lombard speech.

II. METHODS

In this study, two types of analyses are performed. In the first investigation, naturally produced Lombard speech was evaluated with CI users to determine the perceptual effects using two alternate simulated noisy environments with varying SNR. In the second investigation, an acoustic LE perturbation algorithm was developed to evaluate intelligibility of CI users in the presence of large crowd noise with the hypothesis that such modifications could increase intelligibility for CI users in the presence of noise. Perceptual benefits are discussed for both investigations.

A. Analysis of natural Lombard speech in noise

1. Subjects

Two normal hearing (NH) speakers (one male and one female) were recruited from the University of Texas at Dallas to develop a small Lombard speech corpus. Both normal hearing subjects were native speakers of American English without any reported history of speech, language, or hearing problems. Five cochlear implant users (two male, three female) were recruited for this study and paid for their participation. Demographical information is presented in Table I. All CI users were English-speaking, post-lingually deafened adults implanted with Nucleus devices from Cochlear Ltd. The Advanced Combination Encoder (ACE) strategy was used routinely in the participant’s commercial processor. All CI users who participated in the analysis of natural Lombard speech in noise also participated in the perturbation investigation.

2. Stimuli

To develop the testing battery, each NH speaker was exposed to a single noise type, large crowd noise through worn open-air headphones at three presentation levels, 70,
80, and 90 dB SPL. In order to eliminate the occlusion effect, open-ear headphones were used to present noise and provide audio feedback of the speaker to enable LE in recording mode that resulted in noise-free LE recordings. Large crowd noise (LCN) samples were recorded in the UTDallas Student Center using a LENA device (LENA Foundation, 2014). In a study by Hansen and Varadarajan, different variations or “flavors” of LE speech were produced using exposure to varying noise types and levels [i.e., nine flavors of LE were established (Hansen and Varadarajan, 2009)]. In this current study, LCN was chosen as the noise exposure type to produce Lombard speech as it influences the majority of the speech spectrum (0.5–4 KHz) different than that of babble noise which can increase the difficulty of speech perception with decreasing SNR (Krishnamurthy and Hansen, 2009) and to represent naturalistic noisy listening environments for CI users. LCN was also used to develop the noisy conditions used in the CI subjective evaluation representing matched cases, whereas evaluating with a different noise exposure type would represent an unmatched case (Lu and Cooke, 2008). A total of 660 sentences from the AzBio database were read by NH speakers in a sound booth in a conversation-style setup (Spahr et al., 2012). Both speakers were placed on either side of a table approximately 1 m away. Sentence tokens were displayed on an LCD screen in a neutral position between the speakers at eye level so the speakers could direct the prompted sentence to their NH partner in order to encourage prompted voice engagement. Spoken responses were recorded using a close-talk headset microphone worn to capture noise-free Lombard speech (i.e., all recordings were void of noise to representing clean LE speech). The distance of the microphone and the speaker was held 5 cm apart. The speakers were able to repeat sentences, initiate breaks, and adjust for comfort as needed. For the quiet condition, speakers read sentences without noise presented in the open-ear headphones. For CI subject evaluation, clean sentences were mixed with 10 and 15 dB SNR LCN for each of the three LE speaking conditions (e.g., LOM70, LOM80, and LOM90, representing a 70, 80, and 90 dB SPL noise exposure). A total of 12 conditions were generated and used as the testing battery in this phase of the study.

### 3. Procedure

A listening test was conducted within an anechoic sound booth with CI users. Sentence tokens were presented at 60 dB SPL through loud speakers approximately 1 m away from the subject. The 12 conditions consisted of three varying noise conditions: quiet, 10, and 15 dB SNR, and four varying Lombard speaking styles: quiet, 70, 80, 90 dB SPL noise exposure totaling 240 sentences for Analysis A excluding training. For each condition, ten sentences were produced from the male speaker and ten sentences were produced from the female speaker. Conditions were scored for words correct to determine human speech recognition. All sentence tokens were randomized for noise condition and Lombard speaking style. A short training session of five sentences was played for each subject to provide familiarization of each condition. No repetitions were allowed during the entirety of the experiment to avoid speech recognition repetition effects. The duration of the testing session was 2 h with intermittent breaks for the perceptual analysis.

### 4. Statistical analysis

To determine the effect of Lombard speaking style, a repeated-measures, two-way analysis of variance (ANOVA) was performed using SI scores across each Lombard speaking condition. Dunnett’s multiple comparisons test was used to identify individual differences with a 95% confidence level. Statistical analysis was evaluated using GraphPad Prism (GraphPad Software Inc., 2019, Prism 8 for Windows, Version 8.2.0, San Diego, CA).

#### B. Analysis of simulated neutral perturbed Lombard speech in noise

1. **Signal processing—Perturbation algorithms**

   A prior analysis of LE specifically for CI users indicated statistically meaningful changes in speech production of neutral speech and Lombard-produced speech in both the frequency and time domain (Lee et al., 2015; Lee et al., 2017). The authors determined LE can be produced by both normal hearing and CI users (Lee et al., 2015; Lee et al., 2017). A three stage signal processing approach was used to transform neutral speech to Lombard-style speech by (1) temporal amplification, (2) spectral contour modification, and (3) sentence duration modification. This three-step process was developed using the statistical source generator theory proposed by Hansen and Cairns (Hansen, 1994; Hansen and Cairns, 1995). Source generator theory defines variations of neutral to Lombard speech as a statistical path or model of the speech production space based on the noise exposure. Front-end speech processing algorithms primarily used in automatic speech recognition (ASR) and speaker ID

---

TABLE I. Demographical information of CI users (N = 5) participating in both perceptual analysis (A) and perturbation investigation (B).

| Subject | Sex | Age (yr) | Hearing loss (yr) | Implantation (yr) | Etiology of hearing loss | Implant No. (ear tested) | Coding strategy |
|---------|-----|----------|-------------------|-------------------|-------------------------|-------------------------|-----------------|
| S1      | M   | 64       | 6                 | 2                 | Noise                   | Bilateral (L)           | ACE             |
| S2      | M   | 78       | 38                | 1                 | Noise                   | Bilateral (R)           | SPEAK           |
| S3      | M   | 70       | 17                | 7                 | Hereditary              | Unilateral (L)          | ACE             |
| S4      | F   | 69       | 13                | 7                 | Hereditary              | Bilateral (R)           | ACE             |
| S5      | F   | 60       | 30                | 5                 | Hereditary              | Bilateral (R)           | ACE             |
(SID) systems, were used to compensate or reduce the LE known to degrade system performance (Boril and Hansen, 2010; Hansen and Cairns, 1995; Hansen, 1994; Hansen and Varadarajan, 2009; Hansen et al., 2011; Kelly and Hansen, 2016; Saleem et al., 2015). For the reverse case, the variations defined from the source generator theory were used in the development of a perturbation algorithm to modify neutral speech into the Lombard speech domain to improve acoustic models for ASR applications (Bou-Ghazale and Hansen, 1996, 1997, 1998, 2000). Modification of speaking style in the study from Bou-Ghazale and Hansen utilized hidden Markov models (HMMs) to transform the duration, pitch, and spectral slope of neutral speech to that of Lombard speech and observed notable improvements of Lombard classified speech in the ASR domain (Bou-Ghazale and Hansen, 1996). Speech modification studies sought to generate Lombard-like speech by altering fundamental acoustic and temporal components of neutral clean speech and correlate changes to from listening experiments with normal or typical hearing users (Cooke et al., 2013; Cooke et al., 2014; Lu and Cooke, 2009; Niederjohn and Grotelueschen, 1976; Scheper et al., 2013; Jokinen et al., 2016; Zorila et al., 2012). Outside from the work of Lee and colleagues, the perception of the LE to CI listeners and the differences of the fundamental spectral-temporal characteristics due to CI-specific signal processing has not been studied (Lee et al., 2015; Lee et al., 2017). Thus, it is hypothesized that the fundamental characteristics of LE discovered from listening experiments with typical hearing individuals can also be leveraged within the signal processing of CIs to provide a natural, physiologically inspired, front-end processing algorithm to migrate speech to Lombard speech as means to improve intelligibility in listening situations with background noise. The UT-Scope (Speech under Cognitive and Physical Stress and Emotion) database was used to model acoustic parameters of speech in noise to develop the spectral mismatch between Lombard and neutral speech (Ikeno et al., 2007). This database includes closed-talk microphone recordings of 59 speakers reading 20 TIMIT sentences with large crowd noise exposure at 80 dB SPL through open-ear headphones within a sound booth (Garofolo et al., 1993; Ikeno et al., 2007). Similar setup was used to collect baseline quiet conditions. The total number of sentence tokens used for variation/parametric modeling was 2360 sentences (2 conditions, 20 sentences each, 59 speakers). Lombard speech has been shown to demonstrate a spectral flattening of speech and thus, previous modification strategies provided related boosts and attenuation within the frequency ranges associated with and without speech as in Zorila et al. (2012) or by linear regression as in Lu and Cooke (2009) (Dreher and O’Neil, 1957; Godoy and Stylianou, 2013; Godoy et al., 2014; Junqua, 1992; Lee et al., 2015, 2017; Lu and Cooke, 2009). A time-invariant spectral mismatch filter as opposed to a low-pass filter was used to account for Lombard-like spectral differences as done similarly by Godoy and colleagues who modeled differences in envelopes of neutral and Lombard speech (Godoy and Stylianou, 2013).

The three-stage perturbation algorithm demonstrated in Fig. 1 was processed in an offline manner before presentation to CI users. First, highly intelligible segments of the neutral sentence token were identified using the cochlear-scaled entropy estimate (CSE) (Stilp and Kluender, 2010). It should be noted the CSE estimate identifies vowel-consonant boundaries as well as vowels and consonant individually known to predict intelligibility (Kewley-Pott et al., 2007). Euclidean distances were calculated from adjacent 16 μs segments passed through a 16 band-pass Gammatone filterbank (Patterson et al., 1987). The average Euclidean distance over five successive segments (80 ms) was used to classify the segment’s overall CSE measure. The output of the CSE estimate classifies speech either as “high” or “low” entropy segments based on a proportional coefficient, p. This proportional coefficient represents a p percentage of the speech utterance/sentence below a particular threshold. In this study, p is set to 0.6, which is representative of the segment contributing 40% of speech above this threshold. Speech segments meeting the p threshold constraint were selected for amplification. Signal power of the speech-only utterances were averaged over the sentence for the neutral and Lombard token using PRAAT (Boersma, 2002). The amplification ratio was calculated as the power ratio between the neutral and Lombard tokens. This amplification ratio was used as a scaling factor in the time domain for those sentences meeting the p threshold calculated via CSE measure. Figure 2 illustrates the high-entropy CSE decision in an example TIMIT sentence. Amplification was limited
to 50% to maintain speech segment quantity and integrity. Amplification of speech segments was done within the context of the time domain.

Second, the sentence token was passed through a spectral shaping, time-invariant filter modeled after the speech variations of Lombard-produced speech from the UT-Scope database as shown in Fig. 3 and as discussed previously. Spectral energy of the neutral sentence token was estimated through frame-by-frame processing using a second-order filter from 32 ms segments with 50% frame overlap. The spectral shaping mismatch filter calculates the production differences between the contour of neutral and Lombard speech for the same text context speech utterances. The mismatch was calculated as the difference between the spectral energy of the sentence token and the Lombard-produced speech, redistributing energy across the frequency domain according to the modeled data (Ikeno et al., 2007; Godoy and Stylianou, 2013). Overall power of the sentence token was maintained before and after frequency-based processing. In this modification step, high-frequency regions of speech were increased to develop a flattened spectrum.

Third, durational modifications where applied using the time-domain-pitch-synchronous-overlap-and-add (TD-PSOLA) method. The durational lengths of Lombard speech were modeled from the UT-Scope database (Ikeno et al., 2007) and used to develop a uniform time stretching ratio of the neutral sentence token duration to the corresponding Lombard token. The same frame-by-frame processing parameters were used as in the spectral contour modification. To apply uniform time-stretching ratio to lengthen the sentence duration, speech frames were repeated via TD-PSOLA in order to provide the listener with the greatest possible chance of correctly hearing the speech segment (Bou-Ghazale and Hansen, 1996, 2000; Cooke et al., 2014; Godoy and Stylianou, 2013). Uniform time-stretching ratios were multiplied by the duration of the neutral sentence and root-mean-square normalized to match the original neutral token. The TD-PSOLA scaling ratio was limited to 2 (i.e., the perturbed sentence could not exceed twice the duration of the neutral sentence token).

2. Subjects

Five CI users were recruited and paid for their participation in this pilot study. All participants were native speakers of English, post-lingually deafened, and had more than 6 months of experience with their clinical processors. CI user eligibility of this study was limited to implants manufactured by Cochlear Corp. using a CIS strategy, with ACE routinely used in their clinical processor. All CI subjects in this study participated in the perceptual analysis and the perturbation investigation.

3. Stimuli

To generate Lombard-Effect-perturbed speech, sentences from the AzBio database were processed through the three-stage perturbation algorithm (see Fig. 1) in an offline manner using MATLAB and PRAAT (Boersma, 2002). Perturbed sentences were used in the control condition (quiet) and sentences were mixed with large-crowd-noise at two different SNRs: 15 and 10 dB SPL for the noise conditions. To determine the effect of each stage within the perturbation algorithm, additional AzBio sentences were perturbed from each stage of the perturbation strategy alone: temporal amplification (amplification-modified), spectral-mismatch filtering (spectrum-modified), or uniform-time-stretching (duration-modified). The

**FIG. 2.** (Color online) (a) Spectrogram of a neutral-spoken sentence from the TIMIT database; (b) temporal modification using the cochlea-scaled entropy (CSE) decision to identify highly intelligible segments of speech (Stilp and Kluender, 2010). The blue plot demonstrates the weight output of the CSE and the red plot demonstrates the amplification of the high-entropy segments. No amplification was performed for low-entropy segments, i.e., when TF is 1, the segment is amplified, when TF is 0, the segment is not amplified.

**FIG. 3.** (Color online) Average frequency responses of the Lombard (80 dB SPL LCN) and neutral sentence tokens from the UT-SCOPE database. The frequency response of the Lombard sentence tokens were used to generate the spectral shaping filter of the LE perturbation algorithm.
18 total conditions consisted of three varying noise conditions: quiet, 10, and 15 dB SNR LCN, five varying perturbation conditions: no perturbation (neutral), amplification-modified, spectrum-modified, duration-modified, and three-stage perturbation strategy, and one natural Lombard speaking condition (LOM90, same processing condition in Analysis A). A total of 360 sentences tokens were used for Analysis B excluding training. Each condition was evaluated from 20 AzBio sentences. The speech battery was presented to the CI user in a sound booth through a loudspeaker setup at 60 dB presentation level. Each subject was evaluated using the number of words correct per condition. CI users were not allowed to listen to the sentence more than once. Number of words correct was scored in the same way as Analysis A.

4. Statistical analysis

To determine the effects of noise condition and artificial modification, the same statistical analysis performed for Analysis A is duplicated. A repeated-measures, two-way ANOVA was performed on the percentage point improvement from neutral speech. Multiple comparisons using Dunnett’s post hoc test was used to determine the effects of each perturbation component, natural Lombard speech (LOM90) from Analysis A, and the LE-perturbation algorithm (combination of all three modification types). Statistical analysis was determined at a 95% confidence level evaluated using GraphPad Prism (GraphPad Software Inc., 2019, Prism 8 for Windows, Version 8.2.0, San Diego, CA).

III. RESULTS

A. Perceptual effect of natural Lombard speech in noise

Average intelligibility scores from natural Lombard evaluation with 5 CI listeners are illustrated in Fig. 4. A significant effect of Lombard speaking style (F[3,12] = 5.019, \( p < 0.02 \)) and noise (F[2,8] = 52.35, \( p < 0.0001 \)) was observed while the interaction (F[6,24] = 0.4766, \( p > 0.05 \)) was not significant from a repeated-measures, two-way ANOVA. As the Lombard speaking style increases with respect to decreasing SNR, an increase in average intelligibility was observed for both quiet and 10 dB SNR conditions. Baseline performance of CI listeners under anechoic quiet condition was 67.5%, which decreased to 38.1% in 15 dB SNR, and 25.4% in 10 dB SNR. In general, the three Lombard conditions produced an average intelligibility restoration of +6.95% compared to baseline. Comparing intelligibility in the noise conditions to the strongest LE condition (90 dB SPL), improvements of +7.6%, +8.4%, and +13.2% were observed in quiet, 15 and 10 dB SNR conditions, respectively. Significant improvement was noted only for the highest Lombard speaking style (LOM90) compared to the neutral baseline (\( p < 0.01 \)).

B. Perceptual effects of LE modified neutral speech in noise

Figure 5 demonstrates the average percentage point improvement or decrement from the neutral baseline condition for the modification approaches in addition to natural Lombard speech (LOM90). With the exception of the quiet condition, the approach yielding the largest benefits was observed with the full, three-step LE perturbation algorithm as compared natural Lombard (LOM90) as well as each of the individual components. LE perturbation strategy resulted in SI gains of +12.8% percentage points for 15 dB SNR and +16.8% percentage points for 10 dB SNR from the neutral baseline. Results from a two-way ANOVA revealed significant effects of LE modification strategy (F[4,16] = 3.29, \( p < 0.02 \)) and the interaction (F[8,32] = 3.105, \( p < 0.02 \)) of modification and noise, but not for the noise condition (F[2,8] = 1.063, \( p > 0.05 \)). Across all three conditions, intelligibility increased on average of +8.1% using the three-stage LE perturbation algorithm. Compared to the average of all three natural Lombard speaking styles (LOM70, LOM80, LOM90) from Analysis A, the LE perturbation demonstrated increased intelligibility by +14.5% and +8.5% percentage points for the noisy conditions, respectively.

Comparison of SI gains from the neutral baseline of each component within the LE perturbation strategy was analyzed and compared across the two noise conditions (see Fig. 5). Performance of CI listeners from LE amplification-only, LE spectrum-only, and LE duration-only resulted in lower intelligibility than the perturbation strategy as a combination of the three individual LE components. From individual acoustic modifications, performance of the LE spectral mismatch filter outperformed the LE duration-only and LE amplification-only. LE perturbation strategy resulted in significant improvements compared to amplification-modified (\( p < 0.01 \)), duration-modified (\( p < 0.02 \)), and natural Lombard (LOM90) (\( p < 0.05 \)) for 15 dB SNR from a
For the 10 dB SNR condition, the LE perturbation algorithm demonstrated significant improvements against amplification-modified ($p < 0.0001$), spectrum-modified ($p < 0.001$), and duration-modified ($p < 0.001$) approaches.

IV. DISCUSSION

Perceptual effects of natural Lombard speaking style with CI listeners was demonstrated by evaluating SI in two modalities: (1) multiple levels of noise exposure resulting in various levels or speaking styles of LE speech and (2) multiple levels of additive large-crowd-noise. The latter speech-in-noise task compared CI performance from noise-free LE speech without additive noise to noise-free LE speech with additive LCN. Results indicate SI significant improvements only for the 10 dB SNR LCN ($p < 0.001$) condition in the natural LE in Analysis A. Larger improvements were observed for the high noise exposure producing the greatest Lombard speaking style (LOM90). Speech produced using LE in the simulated noise conditions demonstrated the perceptual benefit of intelligibility for CI users in challenging acoustic environments, but not significantly different than neutral speech. These results are differ from previous LE studies for NH users (Lu and Cooke, 2008, 2009; Cooke et al., 2013) where significant benefits were achieved.

Previous studies demonstrated speech spoken in quiet environments is found to be less intelligible than Lombard speech (Dreher and O’Neill, 1957; Junqua, 1992; Lu and Cooke, 2008; Pickett, 1956; Pittman and Wiley, 2001; Summers et al., 1988). Analysis of the acoustic and phonetic changes between LE speech produced in quiet compared to LE speech produced in noise (additive) may contribute to the SI improvement (Hansen, 1988; Junqua, 1992; Lee et al., 2017; Lu and Cooke, 2008). Pickett noted an increase in intelligibility gain as the environmental noise levels became more severe (Pickett, 1956). Lower speech recognition performance of extreme vocal effort including physical shouting has also been reported (Picheny et al., 1986; Rostolland and Parant, 1973). These findings assist in determining the range of vocal effort as a function of SI. This range is not only useful to define for all listeners, but provides rationale for possible perceptual benefits of LE speech in every-day noisy conditions for CI users.

Acoustic modifications of speech such as flattened spectral tilt with increased high frequency content, adjustments in formant frequencies, and other phonetic durational changes have been associated with LE (Bou-Ghazale and Hansen, 1996, 1997, 1998, 2000; Hansen, 1988, 1989; Hansen and Bria, 1990; Lee et al., 2015; Lee et al., 2017). These acoustic modifications can be used for both for NH and CI listeners. One way to visualize these modifications as verification in LE speaking styles in this study is to inspect electrical stimulus patterns otherwise known as electrodograms. Similar to a spectrogram, electrodograms provide a time-frequency representation of electrical current sent to the intracochlear electrode array. Figure 6 demonstrates four electrodograms of the sentence, “Basketball can be an entertaining sport,” from the UT-SCOPE database (Ikeno et al., 2007) using a 22-band
ACE-processing CI signal processing strategy (Vandali et al., 2000). There exist three notable patterns of electrical activity from comparing LE condition with the neutral baseline. LE stimuli in Figs. 6(b) and 6(d), in noise and in quiet conditions, provides more electrical activity in high frequency regions (electrodes 1–12) which is correlated to greater spectral energy. For reference, the center frequency of electrode 11 is approximately 1700 Hz according to frequency allocations provided from Cochlear Ltd. Second formant transition, the third formant frequency, and consonants located in this region may therefore be emphasized further in LE to provide contrast between speech and noise. The impact of large crowd noise in Figs. 6(c) and 6(d) appears to distort lower frequencies (electrodes 13–22) more than higher frequencies. Last, the increase in energy in high frequency regions results in an overall flattened frequency spectrum over time compared to the multi-peak spectrum for neutral speech. These contributions demonstrated in Fig. 6 indicate possible perceptual benefits of natural acoustic modifications employed by LE.

Individual contributions to SI, however, did not outperform the LE perturbation algorithm. It is suggested that listeners have an inherent trained context for noise exposure, in that the combination of multiple speech production modifications due to LE must be present before the listener can decode the perturbed/modified neutral speech as having LE intelligibility benefits. It may also be suggested that CI listeners lack of exposure or inherent training to perceive noise-free Lombard speech outside the context of the noisy background environment explaining the lack of significant differences in the natural LE listening experiment in analysis A. When evaluating LE spectrally modified speech, the amplification or power of high-frequency components may be at the expense of low-frequency components (Jokinen et al., 2016; Niederjohn and Groteleueschen, 1976; Schepker et al., 2013). Another study indicated an increase in consonant energy at the expense of vowel energy.
called “clear speech” (Bradlow et al., 1994, 1997; Hansen, 1994, 1996; Spahr et al., 2002; Zeng et al., 2002; Zorila, 2012). These two functions exist in virtually all current commercial sound processors today, differing only based on manufacturer. In this study, a pre-processing perturbation strategy has been suggested as for integration within the signal processing pipeline as a pre-processing component. In its present form, the pre-processing algorithm is not void of limitations. Fundamental frequency was not included in the perturbation algorithm due to poor pitch perception of CI users regarded from temporal fine structure degradation (Moore, 2008). Real-time implications were not considered during the development of this study, due to durational-modifications through uniform time stretching which required processing the entire sentence token before modification. It should also be noted that this investigation used noise-free recordings of Lombard speech for both analyses (natural LE and artificial LE). Future work should compare the performance of the LE-perturbation from an original noisy speech token as it is more indicative of real-world listening situations for CI users. Behind-the-ear (BTE) microphones on CI clinical processors may or may not use multi-channel or microphone arrays preventing the LE-perturbation strategy in its current state from receiving clean speech. The development of the LE spectral mismatch filter only consisted of variation changes in the frequency domain from speakers exposed to only one noise level and type, 80 dB SPL. A future investigation of lower and upper bounds of intelligibility gains can be explored to determine alternative filter designs yielding universal gains across the majority of CI users. Additionally, future investigations can be done to analyze varying LE spectral mismatch filters based on different noise types than the one considered here based on large-crowd noise.

V. CONCLUSION

This study has investigated how natural and artificially produced/perturbed LE speech positively effects speech intelligibility in noisy conditions for CI users. Through the perceptual analysis of varying LE speaking styles (increasing noise exposure to produce various flavors of natural LE speech), restoration of intelligibility in noisy conditions was observed. Mimicking LE through a perturbation algorithm incorporated three alternative modification techniques based on (i) duration modification, (ii) temporal amplification of highly intelligible segments, and (iii) spectral mismatch filtering. Together, these components contributed to significant improvements in intelligibility, on average a +14% percentage point improvement in large-crowd noise conditions. The LE modifications evaluated in this study yielded implications for future signal processing of cochlear implants. Pre-processing algorithms are commonly used for speech enhancement of noise suppression approaches, but perturbation or synthesized speech has not
yet breached the surface of commercial CI sound processors. Results from this study can serve as the rationale for developing acoustic, temporal, and spectral modifications to existing processing paradigms.

Assmann, P. F., and Summerfield, Q. (2004). “The perception of speech under adverse conditions,” in Handbook of Audiology Research, edited by W. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 231–308.

Boersma, P. (2002). “Praat, a system for doing phonetics by computer,” Glot Int. 5, 341–345.

Boril, H., and Hansen, J. H. L. (2010). “Unsupervised equalization of Lombard Effect for speech recognition in noisy adverse environments,” IEEE Trans. Audio Speech Lang. Proc. 18(6), 1379–1393.

Bou-Ghazale, S. E., and Hansen, J. H. L. (1996). “Generating stressed speech from neutral speech using a modified CELP vocoder,” Speech Commun. 20(1-2), 93–110.

Bou-Ghazale, S. E., and Hansen, J. H. L. (1997). “A novel training approach for improving speech recognition under adverse stressful conditions,” in Proceedings from the European Conference on Speech Communication and Technology (EUROSPEECH), Rhodes, Greece, pp. 2587–2590.

Bou-Ghazale, S. E., and Hansen, J. H. L. (1998). “Speech feature modeling for robust stressed speech recognition,” in Proceedings from the international Conference on Spoken Language Processing (ICSLP), Sydney, Australia, p. 918.

Bou-Ghazale, S. E., and Hansen, J. H. L. (2000). “A comparative study of traditional and newly proposed features for recognition of speech under stress,” IEEE Trans. Speech Audio Proc. 8(4), 429–442.

Bradlow, A. R., Kraus, N., and Hayes, E. (2003). “Speaking clearly for children with learning disabilities: Sentence perception in noise,” J. Speech Lang. Hear. Res. 46(6), 80–97.

Cook, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauerb, B., and Tang, Y. (2013). “Evaluating the intelligibility benefit of speech modifications in known noise conditions,” Speech Commun. 55, 572–585.

Cook, M., Mayo, C., and Villegas, J. (2014). “The contribution of duration and spectral changes to the Lombard speech intelligibility benefit,” J. Acoust. Soc. Am. 135(5), 874–883.

Deng, L., Ramsay, G., and Sun, D. (1997). “Production models as a structural basis for automatic speech recognition,” Speech Commun. 22(2-3), 93–111.

Docolo, S., Kellermann, W., Makino, S., and Nordholm, S. E. (2015). “Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones,” IEEE Sign. Proc. Mag. 32(2), 18–30.

Donaldson, G. S., and Allen, S. L. (2003). “Effects of presentation level on phoneme and sentence recognition in quiet by cochlear implant listeners,” Ear Hear. 24(5), 392–405.

Dorman, M., and Spahr, A. (2006). “Speech perception by adults with multichannel cochlear implants,” in Cochlear Implants, 2nd ed. (Thieme Medical Publishers, New York), pp. 193–204.

Dorman, M. F., Hannley, M. T., Dankowski, K., Smith, L., and McCandless, G. (1989). “Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant,” Ear Hear. 10(1), 44–49.

Dowell, R. C., Mecklenburg, D. J., and Clark, G. M. (1986). “Speech recognition for 40 patients receiving multichannel cochlear implants,” Arch. Otolaryngol.—Head Neck Surg. 112(10), 1054–1059.

Dreher, J. J., and O’Neill, J. (1957). “Effects of ambient noise on speaker intelligibility for words and phrases,” J. Acoust. Soc. Am. 29, 1320–1323.

Duddington, D. K. (1980). “Speech discrimination in deaf subjects with cochlear implants,” J. Acoust. Soc. Am. 68(3), 885–891.

Ferguson, S. H., and Kewley-Port, D. (2002). “Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. 112(1), 259–271.

Fu, Q.-J., Shannon, R. V., and Wang, X. (1998). “Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,” J. Acoust. Soc. Am. 104(6), 3586–3596.

Garner, M., Henrich, N., and Dubois, D. (2010). “Influence of sound immersion and communicative interaction on the Lombard effect,” J. Speech Lang. Hear. Res. 53, 588–608.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., and Zue, V. (1993). “TIMIT acoustic-phonetic continuous speech corpus,” in Linguistic Data Consortium, Philadelphia, PA, p. 33.

Godoy, E., Koutsogiannaki, M., and Stylianou, Y. (2014). “Approaching speech intelligibility enhancement with inspiration from Lombard and clear speaking styles,” Comput. Speech Lang. 28(2), 629–647.

Godoy, E., and Stylianou, Y. (2012). “Unsupervised acoustic analyses of normal and Lombard speech, with spectral envelope transformation to improve intelligibility,” in Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH), Portland, OR, pp. 1472–1475.

Hansen, J. H. L. (2013). “Increasing speech intelligibility via spectral shaping with frequency warping and dynamic range compression plus transient enhancement,” in Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH), Lyon, France, pp. 3572–3576.

Hansen, J. H. L. (1988). “Analysis and compensation of stressed and noisy speech with application to robust automatic recognition,” Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA.

Hansen, J. H. L. (1989). “Evaluation of acoustic correlates of speech under stress for robust speech recognition,” in Proceedings of the IEEE Fifteenth Annual Northeast Bioengineering Conference, Boston, MA, pp. 31–32.

Hansen, J. H. L. (1994). “Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect,” IEEE Trans. Speech Audio Proc. 2, 598–614.

Hansen, J. H. L. (1996). “Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition,” Speech Commun. 20(2), 151–170.

Hansen, J. H. L., and Bria, O. (1990). “Lombard effect compensation for robust automatic speech recognition in noise,” in Proceedings of the International Conference on Spoken Language Processing (ICSLP-90), Kobe, Japan, pp. 1125–1128.

Hansen, J. H. L., and Cairns, D. A. (1995). “Icarus: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments,” Speech Commun. 16(4), 391–422.

Hansen, J. H. L., Sangwan, A., and Kim, W. (2011). “Speech processing for robust speaker recognition: Advancements for speech under stress, Lombard Effect, and Emotion,” in Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism (Springer-Verlag, New York), Chap. 5, pp. 103–123.

Hansen, J. H. L., and Varadarajan, V. (2009). “Analysis and compensation of Lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition,” IEEE Trans. Audio Speech Lang. Proc. 17(2), 366–378.

Hansen, J. H. L., and Womack, B. “Feature analysis and neural network based classification of speech under stress,” IEEE Trans. Speech Audio Proc. 4(4), 307–313 (1996).

Hazen, V., and Markham, D. (2004). “Acoustic-phonetic correlates of talker intelligibility for adults and children,” J. Acoust. Soc. Am. 116(5), 3108–3118.

Hazarati, O., Sadjadi, S. O., and Hansen, J. H. L. (2014). “Robust and efficient environment detection for adaptive speech enhancement in cochlear implants,” in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 900–904.

House, A. S., Williams, C. E., Hecker, M. H., and Kryter, K. D. (1965). “Articulation-testing methods: Consonantal differentiation with a closed-response set,” J. Acoust. Soc. Am. 37, 158–166.

Hu, Y., and Loizou, P. C. (2007). “A comparative intelligibility study of single-microphone noise reduction algorithms,” J. Acoust. Soc. Am. 122(3), 1777–1786.

Hu, Y., and Loizou, P. C. (2010). “Environment-specific noise suppression for improved speech intelligibility by cochlear implant users,” J. Acoust. Soc. Am. 127(6), 3689–3695.

Ikeno, A., Varadarajan, V., Patil, S., and Hansen, J. H. L. (2007). “UT-Scope: Speech under Lombard effect and cognitive stress,” in Proceedings of IEEE Aerospace Conference, Big Sky, MT, pp. 1–7.

Jaitly, N., and Hinton, G. E. (2013). “Vocal tract length perturbation (VTLP) improves speech recognition,” in Proceedings of ICML Workshop on Deep Learning for Audio, Speech and Language, Vol. 117.
Jokinen, E., Remes, U., and Alku, P. (2016). “The use of read versus conversational Lombard speech in spectral tilt modeling for intelligibility enhancement in near-end noise conditions,” in Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH), San Francisco, CA, pp. 2771–2775.

Junqua, J.-C. (1992). “The Lombard reflex and its role on human listeners and automatic speech recognizers,” J. Acoust. Soc. Am. 93, 510–524.

Kelly, F., and Hansen, J. H. L. (2016). “Evaluation and calibration of Lombard effects in speaker verification,” in IEEE Spoken Language Technology Workshop, San Diego, CA.

Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (2007). “Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing impaired listeners,” J. Acoust. Soc. Am. 122(4), 2365–2375.

Kokkinakis, K., Azimi, B., Hu, Y., and Friedland, D. R. (2012). “Single and multiple microphone noise reduction strategies in cochlear implants,” Trends in Hearing 16(2), 102–116.

Krause, J. C., and Braida, L. D. (2004). “Acoustic properties of naturally produced clear speech at normal speaking rates,” J. Acoust. Soc. Am. 115, 362–378.

Krishnamurthy, N., and Hansen, J. H. L. (2009). “Babble noise: Modeling, analysis, and applications,” IEEE Trans. Audio Speech Lang. Proc. 17(7), 1394–1407.

Lee, J. (2017). “Lombard effect in speech production by cochlear implant users: Analysis assessment and implications,” Ph.D. Dissertation, University of Texas at Dallas, Dallas, TX.

Lee, J., Ali, H., Ziaei, A., and Hansen, J. H. L. (2015). “Analysis of speech and language communication for cochlear implant users in noisy Lombard conditions,” in Proceedings from the IEEE International Conference on Audio and Speech Signal Processing (ICASSP), Brisbane, Australia, pp. 5132–5136.

Lee, J., Ali, H., Ziaei, A., Tobey, E. A., and Hansen, J. H. L. (2017). “The Lombard effect observed in speech produced by cochlear implant users in noisy environment: A naturalistic study,” J. Acoust. Soc. Am. 141(4), 2788–2799.

Loizou, P. C. (2013). Speech Enhancement: Theory and Practice (CRC press, Boca Raton, FL), pp. 69–93.

Loizou, P. C., Lobo, A., and Hu, Y. (2005). “Subspace algorithms for noise reduction in cochlear implants,” J. Acoust. Soc. Am. 118(5), 2791–2793.

Lombard, E. (1911). “Le signe de lelevation de la voix,” Ann. Maladies de l’Oreille Larynx Nez Pharynx 37, 101–119.

Lu, Y., and Cooke, M. (2008). “Speech production modifications produced by competing talkers, babble, and stationary noise,” J. Acoust. Soc. Am. 124, 3261–3275.

Lu, Y., and Cooke, M. (2009). “The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise,” Speech Commun. 51, 1252–1262.

Moore, B. C. (2008). “The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people,” J. Assoc. Res. Otologyng. 9(4), 399–406.

Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (2012). “Measuring speech recognition in children with cochlear implants in a virtual classroom,” J. Speech Lang. Hear. Res. 55, 532–540.

Niederjohn, R., and Grotelueschen, J. (1976). “The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression,” IEEE Trans. Acoust. Speech Sign. Proc. 24(4), 277–282.

O’Shaughnessy, D. (2003). “Interacting with computers by voice: Automatic speech recognition and synthesis,” Proc. IEEE 91(9), 1272–1305.

Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). “An efficient auditory filterbank based on the gammatone function,” in Proceedings of a Meeting of the IOC Speech Group on Auditory Modelling at RSRE, Vol. 2.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (1994). “Intelligibility of conversational and clear speech in noise for listeners with normal and impaired hearing,” J. Acoust. Soc. Am. 95, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). “Speaking clearly for the hard of hearing in acoustic characteristics of clear and conversational speech,” J. Speech Lang. Hear. Res. 29(4), 434–446.

Pickett, J. M. (1956). “Effects of vocal force on the intelligibility of speech sounds,” J. Acoust. Soc. Am. 28, 902–905.