The Lombard effect observed in speech produced by cochlear implant users in noisy environments: A naturalistic study

Jaewook Lee, Hussnain Ali, Ali Ziaei, Emily A. Tobey, and John H. L. Hansen

---

Articles you may be interested in

Analysis of human scream and its impact on text-independent speaker verification
The Journal of the Acoustical Society of America 141, 2957 (2017); 10.1121/1.4979337

Syllable-constituent perception by hearing-aid users: Common factors in quiet and noise
The Journal of the Acoustical Society of America 141, 2933 (2017); 10.1121/1.4979703

Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback
The Journal of the Acoustical Society of America 141, 2758 (2017); 10.1121/1.4981139

Directional asymmetries reveal a universal bias in adult vowel perception
The Journal of the Acoustical Society of America 141, 2857 (2017); 10.1121/1.4981006

Infants adapt to speaking rate differences in word segmentation
The Journal of the Acoustical Society of America 141, 2569 (2017); 10.1121/1.4979704

Adaptation to novel foreign-accented speech and retention of benefit following training: Influence of aging and hearing loss
The Journal of the Acoustical Society of America 141, 2800 (2017); 10.1121/1.4980063

The Lombard effect observed in speech produced by cochlear implant users in noisy environments: A naturalistic study

Jaewook Lee, Hussnain Ali, Ali Ziaei, Emily A. Tobey,[a] and John H. L. Hansen[b]

Center for Robust Speech Systems–Cochlear Implant Lab (CRSS-CIL), Department of Electrical Engineering,
The University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080, USA

The Lombard effect is an involuntary response speakers experience in the presence of noise during voice communication. This phenomenon is known to cause changes in speech production such as an increase in intensity, pitch structure, formant characteristics, etc., for enhanced audibility in noisy environments. Although well studied for normal hearing listeners, the Lombard effect has received little, if any, attention in the field of cochlear implants (CIs). The objective of this study is to analyze speech production of CI users who are postlingually deafened adults with respect to environmental context. A total of six adult CI users were recruited to produce spontaneous speech in various realistic environments. Acoustic-phonetic analysis was then carried out to characterize their speech production in these environments. The Lombard effect was observed in the speech production of all CI users who participated in this study in adverse listening environments. The results indicate that both suprasegmental (e.g., F0, glottal spectral tilt and vocal intensity) and segmental (e.g., F1 for /i/ and /u/) features were altered in such environments. The analysis from this study suggests that modification of speech production of CI users under the Lombard effect may contribute to some degree an intelligible communication in adverse noisy environments.
© 2017 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4979927]

I. INTRODUCTION

A. Speech produced in noisy environments

Changes in speech production resulting from auditory feedback are an important research domain for improving interpersonal as well as human-to-machine communications. For example, in the presence of noise, a speaker experiences a natural phenomenon known as the Lombard effect (Lombard, 1911; Lane and Tranel, 1971; Hansen, 1988; Junqua, 1996; Lu and Cooke, 2008; Garnier *et al.*, 2010). This phenomenon is physiologically realized with altered vocal efforts, such as an increase in vocal intensity, fundamental frequency, glottal spectral slope, or formant structures. The speech production modification help to maintain speech intelligibility during a conversation in challenging listening environments.

The Lombard effect has been widely studied in automatic speech systems, where it is known to degrade the quality of automatic speech recognition (ASR) and speaker identification (SID) systems (Junqua, 1992; Hansen, 1996; Hansen and Varadarajan, 2009; Bořil and Hansen, 2010). Since there are fundamental differences in Lombard speech as compared to neutral speech, Lombard speech greatly contributes to breakdown in speech system performance when systems are modeled with neutral speech. A range of signal processing techniques have been proposed to compensate for the Lombard effect in speech to improve the robustness of the speech systems (Hansen, 1988, 1996; Bou-Ghazale and Hansen, 2000; Bořil and Hansen, 2010). Although well documented for normal hearing listeners as well as automatic speech systems, the Lombard effect has received little, if any, attention for assistive hearing devices, such as cochlear implants (CIs).

B. Background and motivation

A CI is an electronic device that is surgically implanted in the inner ear which directly stimulates the auditory nerve fibers to provide a sensation of sound (Wilson *et al.*, 1991; Loizou, 1999; Zeng *et al.*, 2008). CIs help individuals with profound hearing loss to communicate by providing perceptual benefit as well as allowing auditory feedback in the speech production of CI users (Svirsky *et al.*, 2000; Dorman and Spahr, 2006). The latter of which is important to maintain communication in widely varying naturalistic environments.

Previous studies have considered how restoration of auditory function with a CI may lead to changes in speech production (Hochmair-Desoyer *et al.*, 1981; Kirk and Edgerton, 1983; Svirsky and Tobey, 1991; Svirsky *et al.*, 1992; Matthies *et al.*, 1996; Vick *et al.*, 2001; Lane *et al.*, 2007). Kirk and Edgerton (1983), for example, examined the suprasegmental properties of four postlingually deafened adults who received a House single channel processor. They reported nearly normal fundamental frequencies in speech production of two male and two female postlingually deafened adult recipients of implants. Hochmair-Desoyer *et al.* (1981) also suggested

[a]Also at: School of Behavioral and Brain Sciences, The University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, USA.

[b]Also at: School of Behavioral and Brain Sciences, The University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, USA. Electronic mail: john.hansen@utdallas.edu

improved quality of speech production for adventitiously deaf adults who were implanted with the 3M/Vienna processor. In that study, changes in vowel production and fundamental frequency of speech were observed when compared to subjects' speech characteristics before implantation.

In addition to longitudinal studies, some groups have investigated more precise roles of auditory feedback which affect speech production. Svirsky and Tobey (1991), for example, suggested that auditory feedback plays a calibration role in the control of speech production. According to their research, rapid changes in vowel formant frequencies were observed within a few seconds of turning the speech processor on or off. It was also argued by Svirsky et al. (1992) that many suprasegmental features, including pitch period and vowel duration, demonstrated instantaneous response to the short-time deprivation of auditory feedback. More recently, a number of studies have confirmed the dual role of auditory feedback in speech production of postlingually deafened CI users (Matthies et al., 1996; Vick et al., 2001; Lane et al., 2007). These findings suggest that, auditory feedback is not only used to calibrate the segmental features of speech in the long-term (e.g., pre- and post-implantation), but is also used to regulate suprasegmental features in the short-time domain (e.g., from a few seconds to minutes).

The studies cited above establish that speech production is heavily influenced by any form of auditory feedback. However, changes in speech production when auditory feedback is artificially distorted (e.g., turning processor on/off) does not necessarily provide equivalent scientific understanding about speech production in real communication conditions, such as noisy environments. To our knowledge, no study has examined the effect of background noise on speech production in CI individuals, or demonstrated the Lombard effect in voice communication of CI users.

C. Objectives and methods

The objective of this study was to analyze and model speech production of CI users with respect to environmental noise conditions. In addition, the study aimed to investigate the effect of auditory feedback on speech production in naturalistic acoustic environments. We observe this effect using mobile personal audio recordings from continuous single-session audio streams collected over an individual's daily life. Prior advancements in this domain include the "Prof-Life-Log" longitudinal study at UT-Dallas (Sangwan et al., 2012; Ziaei et al., 2012, 2013) which explored speech communication in naturalistic daily life.

In order to analyze the effect of naturalistic noise on speech production, a total of six postlingually deafened adult CI users participated in this study. They were asked to produce spontaneous speech in various listening environments on a college campus. Analysis of speech production was accomplished using: (i) characteristics of acoustic environments, (ii) evaluation of subject's listening environment, and (iii) acoustic and phonetic properties of speech production in relation to the listening environment. In the first analysis, various approaches were used to characterize real-world environments, e.g., long-term average spectra, modulation spectra,

and noise sound-pressure level (SPL). The second part of the analysis focused on objective metrics to predict speech quality, namely, signal-to-noise ratios with and without the Lombard effect. Last, the parametric variations in vowel, consonant and individual phoneme production were investigated as a function of varying environments. This involved speech SPL, fundamental frequency, glottal spectral tilt, phoneme duration, and formant frequencies. The analyses outlined here will explore relationships of speech production features of CI users upon varying noise/environment types.

II. METHODS

A. UTD-CI-LENA corpus development

In order to investigate the influence of auditory feedback on speech production, a corpus was developed. This corpus included audio streams of CI participants from their daily lives. The LENA (for "Language ENvironment Analysis") device was used for collecting naturalistic audio from CI users (Gilkerson and Richards, 2008; LENA Foundation, 2014). The LENA device is a lightweight compact digital audio recorder that is capable of capturing mono audio data continuously for up to 16 h. The device was worn by each subject, and captured the participant's daily acoustics, including voice communication and interaction with other people, as well as environments (e.g., noise level and type). Figure 1(a) demonstrates how the device was positioned for collecting the audio data using the LENA unit. A cross pack was used to hold the device inside a pocket made of meshed-material for secure and consistent placement. The device was positioned at the center of the chest where it was held stationary with respect to the subject's mouth (approximately 15–20 cm). Such a set-up made it possible for the unit's microphone to detect the acoustic signal more robustly and consistently (across subjects) against environmental noise/reverberation during data collection.

A total of 6 CI speakers (mean age: 65 yr) who were fitted with the Nucleus devices from Cochlear Ltd. participated in the study. All participants were postlingually deaf (lost hearing after the age of 18) and had been regularly using their CI devices for at least four years. Among the participants, five were bilateral, and one was a unilateral CI users. The unilateral CI user wore a hearing aid in the contralateral ear. It should be noted that bilateral listeners are expected to take advantage of head-shadow effect which offers improved sound localization versus a single implant. This in general, provides better speech perception in noise when the target signal is spatially separated from the masker. Detailed biographical information of the CI participants is presented in Table I. In addition, the same number of normal hearing (NH) speakers (mean age: 37 yr) participated in the study as a pair-wise conversational partner. The CI speakers in this research acted as the primary speaker, while the NH listeners served as the secondary speaker/listener. Note that the objective of this study was to analyze the speech production of CI users in different noisy environments. Both CI and NH subjects were native speakers of American English, and included an equal number of male and female participants in each group.

J. Acoust. Soc. Am. **141** (4), April 2017

Lee *et al.* 2789

(a) Set-up for naturalistic audio acquisition

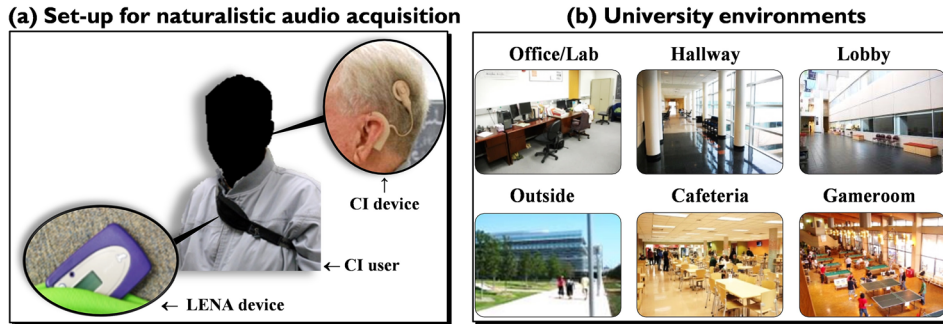**(b) University environments**

FIG. 1. (Color online) Naturalistic data collection from CI subjects: (a) set-up for data acquisition using the LENA device and (b) naturalistic environments on UT-Dallas college campus for data collection.

Naturalistic audio recordings were obtained in six environments on the UT-Dallas college campus. These included (i) office/lab, (ii) hallway, (iii) lobby, (iv) outside on campus, (v) college cafeteria, and (vi) college gameroom, as shown in Fig. 1(b). The locations were chosen to provide a diverse range of noise conditions, for example, type, mixture, and the level varied greatly across environments. Table II summarizes the six naturalistic environments employed for data collection. This consists of (i) general room size, (ii) the number of people typically present during the day, (iii) average SPL, (iv) reverberation time (RT_{60}), and (v) room description. The "people" listing here refers to the typical number of subjects within that room/space. The "average SPL" was determined by calibrating the average noise intensity measured in PRAAT software (Boersma, 2002) and converted to dB SPL scale. The "reverberation time (RT_{60})" refers to the length of time to decay by 60 dB from its initial level of impulsive excitation. The impulsive sound used in the calculation was created by a balloon burst recorded during off-hours in the respective locations.

All audio recordings were collected when on campus population was expected to be consistent with daily conditions. This included data collection on weekdays (Monday to Friday) during normal working hours (10 am to 1 pm) or (1 pm to 4 pm). In order to perform data collection, 3 min of background noise was first recorded prior to the subject's speech production in each location. These noise-only audio segments were then used to assess subject's listening environments in subsequent data analyses. Following the background noise recording, subjects were asked to perform free conversation between each other for 5 min in each location. A list of topics were provided to participants as a suggestion before the test, which included general topics, such as sports, news, weather, movies, etc. The overall data collection

period for each subject was about 2 h. All subjects were informed that they always have the option to pause the audio recording anytime if they might be in a situation where privacy or confidentiality concerns arise during the recording session. However, no interruptions were experienced by any participants.

B. Post-processing

A set of acoustic and orthographic transcription labels were assigned to the collected audio data. The audio streams consisted of two acoustic categories, namely, silence and spontaneous speech in each location. Labeling tasks were first performed by a human annotator based on events in that space. For example, sound events in the office space were different than outside in the public area. In order to produce orthographic transcription, every single isolated utterance (e.g., sentence, phrase, word, and syllable) was first identified manually. Sentence level transcripts were then applied to each identified utterance based on listening to each individual audio file. Additional acoustic labels such as environments (office, hallway, outside, etc.), and speech types (silence, spontaneous) were applied manually to all recordings. In order to reduce inter-labeler variability, only a single annotator, a native speaker of American English, performed all data transcriptions collected from different speakers.

Following the manual labeling tasks, phoneme-level transcription labels were assigned. This task was done automatically by forced speech recognition alignment. Forced alignment is the process of taking the audio file and its orthographic transcription as input to produce word and phone boundary labels. Several other recent studies have successfully used forced alignment as a tool in phonetic research (Lu and Cooke, 2008, 2009; Yu *et al.*, 2014). In this study,

TABLE I. Characteristic information of CI subjects who participated in UTD-CI-LENA corpus development.

Speaker	Gender	Age (yr)	Years of hearing loss	Years implanted	Etiology of hearing loss	Implant ear	Coding strategy
SPK 1	Female	61	56	11	Hereditary	Bilateral	ACE
SPK 2	Female	52	48	5	Hereditary	Bilateral	ACE
SPK 3	Female	61	14	4	Hereditary	Bilateral	ACE
SPK 4	Male	67	12	6	Hereditary	Left only	ACE
SPK 5	Male	81	55	9	Hereditary	Bilateral	ACE
SPK 6	Male	71	18	4	Unknown	Bilateral	ACE

TABLE II. Summary of naturalistic environments used in this study on UT-Dallas campus.

Location	Room size (sq. ft.)	Number of people	Avg. SPL	RT_{60} (sec.)	Room description
Office	450	1–5	43	0.3	An office/lab space with PCs
Hallway	2250	5–10	52	0.6	A long passage with windows
Lobby	11 500	5–50	55	1.8	A 3-story high ceiling space
Outside	Open–air	30–50	61	—	An on-campus outside location
Cafeteria	4500	50–200	65	0.8	A restaurant location with tables
Gameroom	9500	30–75	67	1.3	A public space with billiard tables

an open-source software P2FA was employed for this procedure (Yuan and Liberman, 2008). Following the automatic alignment process, words and phonemes that were shorter than 40 ms in duration, and had fewer than 200 instances were excluded from analysis. A total of approximately 36 000 words including 38 000 vowel and 54 000 consonant nuclei were identified to be analyzed. It is important to note that due to the limited number of contexts, individual phoneme instances should not be regarded as prototypical.

C. Signal processing: Features and metrics

Acoustic characteristics of the background noise were considered by investigating the (i) long-term average spectrum, (ii) average modulation spectrum, (iii) noise SPL, (iv) spectral centroid, and (v) average modulation spectrum energy. It is important to mention here that the noise analysis was carried out on "noise-alone audio segments" which were collected in each environment prior to conversation. None of these samples contained any speech. The long-term average spectrum was obtained by averaging short-time power spectral estimated by the Welch's method. The spectral centroid was calculated based on the average frequency weighted by amplitudes, divided by the sum of the amplitudes. The duration of the analysis window was set at 100 ms with a 50% overlap for each measurement.

We calculated the average modulation spectra to obtain a better understanding of overall room acoustics. The modulation spectrum represents the slowly varying temporal envelope components of signal, thereby providing a degree of acoustic spectral stationarity. Noise samples from each environment were divided into 2-s segments with 1 s time intervals. The modulation spectrum was computed by taking Fourier transform of the Hilbert envelope for each segments. The 0–20 Hz components were then added together across all segments. Finally, the average between 0 and 20 Hz was computed and considered as the average modulation spectrum energy.

Following the noise characteristics, the individual's listening environments were characterized by estimating signal-to-noise ratios (SNRs). In this study, two SNR approaches were employed, which include SNR (i) with neutral speech (SNRN) and (ii) with Lombard speech (SNRL). In these measurements, the assumption is that the *office environment* is a quiet baseline, and speech produced in this location will be *neutral*. SNRN was defined as the energy ratio of neutral speech to noise energy in each environment, which is assumed to be without the Lombard effect as follows:

$$SNRN = 10 \log_{10}\left(\frac{E_{Neutral}}{E_{noise}}\right). \tag{1}$$

And SNRL was calculated from the energy ratio of Lombard speech to the corresponding background noise for each environment as follows:

$$SNRL = 10 \log_{10}\left(\frac{E_{Lombard}}{E_{noise}}\right), \tag{2}$$

where E is the average energy. For these calculations, acoustic boundary detection was employed for separating speech from background noise. The leading and trailing silent intervals derived from each audio stream served as noise segments in each location for computing SNRs. These metrics shows the following two observations: (i) if there is a change in SNR due to noise (SNRN) and (ii) the level to which the decreased SNRs recover due to the Lombard effect is observable and measurable (SNRL).

In addition to noise/environment characteristics, various acoustic and acoustic-phonetic features for speech production were analyzed. These include (i) average speech SPL, (ii) fundamental frequency (F0), (iii) overall spectral tilt, (iv) phoneme duration, (v) first formant frequency (F1) location, and (vi) second formant frequency (F2) location. All measurements except phoneme duration and overall spectral tilt were computed using PRAAT software (Boersma, 2002). The average SPL measurements used here were similar to the metrics employed in the noise analysis, so to ensure connected scales between the measurements when reporting intensity. Phoneme duration was obtained from analysis of the phoneme-level transcripts. Spectral tilt was calculated from the difference between the magnitudes of the first spectral harmonic (H1) and that of the third formant peak (A3), i.e., H1-A3 (Hanson, 1997; Iseli *et al.*, 2007) via PRAAT capability. It should be noted here that the focus was to compute an overall spectral slope. A related study by Hansen (1988) demonstrated changes in glottal spectral slope for various types of speech under stress by averaging individual frame spectral slopes of voiced speech over multiple utterances. Since the focus here is on overall speech content, that approach was not employed.

Acoustic-phonetic features were extracted based on phoneme nuclei boundaries marked by a forced phoneme alignment process. The beginning and ending markers of each phoneme were reduced by 20% to eliminate any transitional effects across phoneme classes. The duration of the analysis window used here for both speech and noise measurements was set to 20 ms with a 10 ms skip rate. After the feature

J. Acoust. Soc. Am. **141** (4), April 2017

Lee *et al.* 2791

LONG-TERM ANALYSIS OF
NATURALISTIC ROOM ACOUSTICS

(a) Long-term average spectra
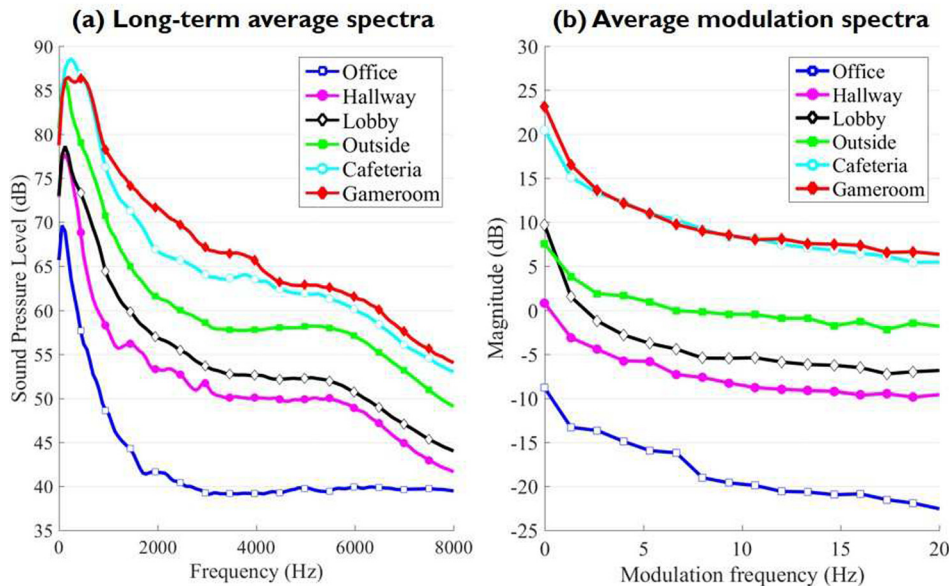
(b) Average modulation spectra

FIG. 2. (Color online) Long-term analysis for six naturalistic environments used in this study on UT-Dallas campus: (a) long-term average spectra and (b) average modulation spectra. Each line corresponds to the average long-term features for each naturalistic environment. The main noise sources were recorded prior to subject's speech production for 3 min in each environment.

extraction, normalization procedures were applied to reduce *speaker-particular* effects from the data (e.g., baseline F0 differences across male and female talkers). This was achieved by scaling each feature to have the same overall level across speakers.

Last, a repeated-measures analysis of variance (ANOVA) was performed to assess the effect of environment type on noise/speech features, and determine statistical significance of differences between speech produced in neutral and Lombard conditions. Subjects were considered as a random (blocked) factors, while environment conditions were used as the main analysis factors. Following the ANOVA, a *post hoc* pairwise comparison test was performed to determine if the noisy conditions were significantly different from the quiet baseline. Bonferrroni adjustment was used to control for family-wise error in the pairwise test. In this study, a difference in means between two or more groups was considered significant if the significance level fell below 5.0% ($p < 0.05$).

III. RESULTS

A. Noise/environment analysis

Prior to any analysis, it is important to understand the characteristics of the acoustic/listening environments. This section offers some level of baseline knowledge regarding each environment's acoustic characteristics as well as how this may relate to speech perception by CI users in these environments.

1. Noise characteristics

Figure 2(a) shows the long-term averaged spectra of various maskers. It can be seen that the office environment has the least spectral impact in terms of overall spectral energy as compared to other noisy environments, which is why it was chosen as a baseline in this study. Spectral energies in general were highly concentrated in the lower frequency range

(<2 kHz) for all environments. When compared to the office baseline, a progressive increase in spectral energies (from hallway to gameroom) was observed in all noisy environments based on the increasing complexity of acoustic space.

This can be better visualized from Figs. 3(a) and 3(b) which present the distribution of average noise SPL and spectral centroid for each environment. The results indicate that both features increased monotonically when switched from office to gameroom. The range of average noise SPL extended from approximately 42 (for office) to 67 dB (for gameroom), and all noisy environments had mean values which were significantly different from the office baseline ($p < 0.05$). Spectral centroid was almost always under 500 Hz for all conditions. With the exception of hallway condition, all noisy environments had mean SPL and spectral centroid values significantly different from the office baseline ($p < 0.05$). Hallway remained almost constant in terms of mean of the spectral centroid ($p > 0.05$).

Figure 2(b) shows the change of the average modulation spectrum as a function of environments. The modulation spectrum energy between the modulation frequencies of 0–20 Hz are presented. All the noise signal has a distinct modulation spectrum with a peak at 0 Hz. All the conditions, with the exception of office environment, had a similar modulation pattern. The modulation spectrum energy between 2 and 20 Hz increased when complexity of noise increased (i.e., with an increase in the number of people, noise SPL). This effect can be better visualized from Fig. 3(c), which shows the average modulation spectrum energy of the tested environments. From this figure, we confirmed the increasing tendency in an explicit way. This suggest that such modulation analysis provides an objective measure of stationarity.

2. Signal-to-noise ratios

Figure 4 illustrates the average SNR levels with respect to each environment. For each environment type, the bar on

TEMPORAL-SPECTRAL CHARACTERISTICS
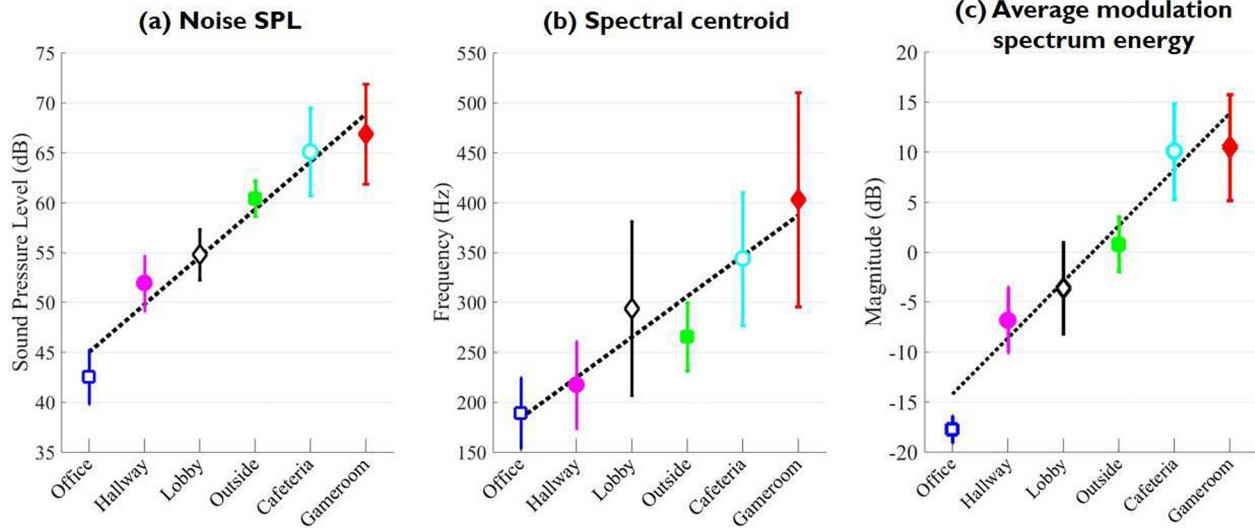## FOR NATURALISTIC NOISE SOURCES

FIG. 3. (Color online) Acoustic characteristics of background noises: average (a) noise SPL, (b) spectral centroid, and (c) average modulation spectrum energy with respect to different environments. While average SPL shows changes in signal strength over time-domain, spectral centroid represents where spectral energy was concentrated in frequency-domain. Average modulation spectrum energy estimate the relative degree of stationarity for the noise signal.

the left (hatched) indicates average SNR without the Lombard effect (SNRN), and the bar on the right (fully colored) indicates average SNR with the Lombard effect (SNRL). With an increase in average SPL of noise at various environments, average SNRN decreased (left-side bars) (baseline office = 34 dB to gameroom = 3 dB), indicating deterioration in speech signal intensity/quality. On the other hand, deterioration in SNRL was not as steep as SNRN (right-side bars) indicating that CI users tend to speak louder in noisy environments to effectively improve the SNR of their speech. This phenomenon is a key characteristic of the Lombard effect. For example, consider the gameroom environment where SNRN was 3 dB. The Lombard effect here helped to increase overall SNR up to 11 dB, thus the corresponding benefit to include the Lombard effect in this environment (gameroom) was + 7.9 dB. The Lombard effect demonstrated here could boost the perceived SNR levels, and thereby facilitate in auditory decoding for the two-way conversations of CI users in noisy conditions.

B. Speech production analysis

In this section, we consider methods for analyzing speech production characteristics as a function of varying environment. Note that in this section, we again established the *office environment* as the quiet baseline (<45 dB SPL), assuming speech production in this location to be *neutral*.

1. Vowel/consonant analysis

For the following analyses, six acoustic features were investigated: (i) vowel SPL, (ii) fundamental frequency F0, (iii) glottal spectral tilt, (iv) vowel duration, (v) consonant SPL, and (vi) consonant duration. While the trajectory of vowel and consonant SPLs indicate the variation in signal strength over time, pitch and spectral tilt deliver the pattern

of shift in spectral cue/energy in the frequency-domain. Here, phoneme duration represents the temporal aspect of the vowel and consonant production.

Figures 5(a) and 5(b) show the average vowel SPL and F0 as a function of varying noise SPL, respectively. The data presented here were averaged across all subjects and all

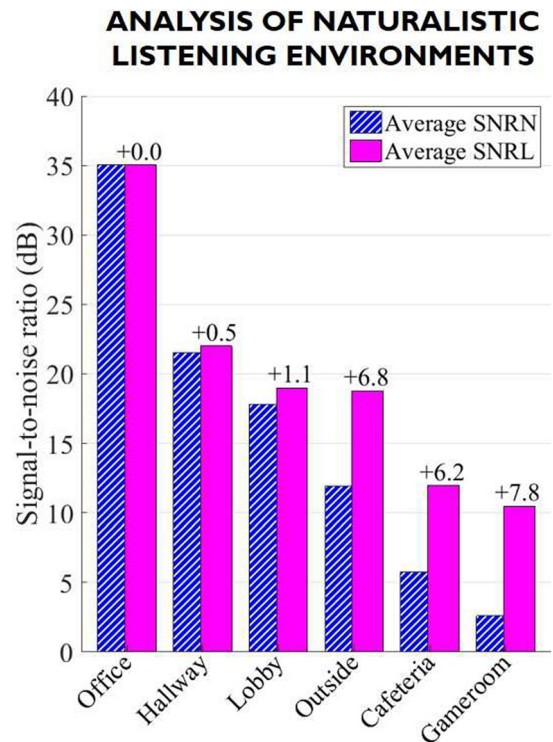## ANALYSIS OF NATURALISTIC
LISTENING ENVIRONMENTS

FIG. 4. (Color online) Evaluation of subject's listening environment using average signal-to-noise ratios as a function of varying environments. Two bars on the left (hatched) and right (fully colored) for each location correspond to the SNR with neutral speech (SNRN) and with Lombard speech (SNRL), respectively.

vowel phonemes. Note that the asterisk marked at respective data points indicate statistical significance as compared to the office baseline ($p < 0.05$). The results indicate that both vowel SPL and F0 varied across conditions. Average values of both features increased significantly for outside, cafeteria, and gameroom environments ($p < 0.05$). However, only little changes occurred in vowel SPL and F0 for hallway and lobby environments ($p > 0.05$). Two groups emerged for both features: a high-value group which was significantly different from office baseline ($p < 0.05$) and a low-value group with no statistically significant difference from baseline ($p > 0.05$). The low value group comprised of the hallway and lobby conditions, while the high value group included outside, cafeteria, and gameroom environments. No significant differences were found between the conditions which belonged to the same group.

Figures 5(c) and 5(d) present the variation of spectral tilt and vowel duration, respectively, with respect to each

environment. Spectral tilt was found to progressively reduce with increasing noise SPL. From the baseline office to the gameroom environment, the mean spectral tilt fell from 19 to 14 dB; with other environments falling within this range. A significant effect of environment type on spectral tilt was observed between gameroom and office baseline ($p < 0.05$). However, no statistical significant differences were found for any combinations of the remaining four conditions (hallway, lobby, outside, cafeteria) ($p > 0.05$). For vowel duration, variations in mean were found in the presence of noise. As shown in Fig. 5(d), average vowel duration decreased progressively with noise complexity; however, it was only significantly different from the baseline condition for the gameroom environment ($p < 0.05$) only. Hallway, lobby, outside, and cafeteria environments did not result in a statistically significant change in the vowel duration as compared to the office baseline.

Figures 5(e) and 5(f) display the variation in consonant SPL and duration, respectively, with respect to noise SPL. In
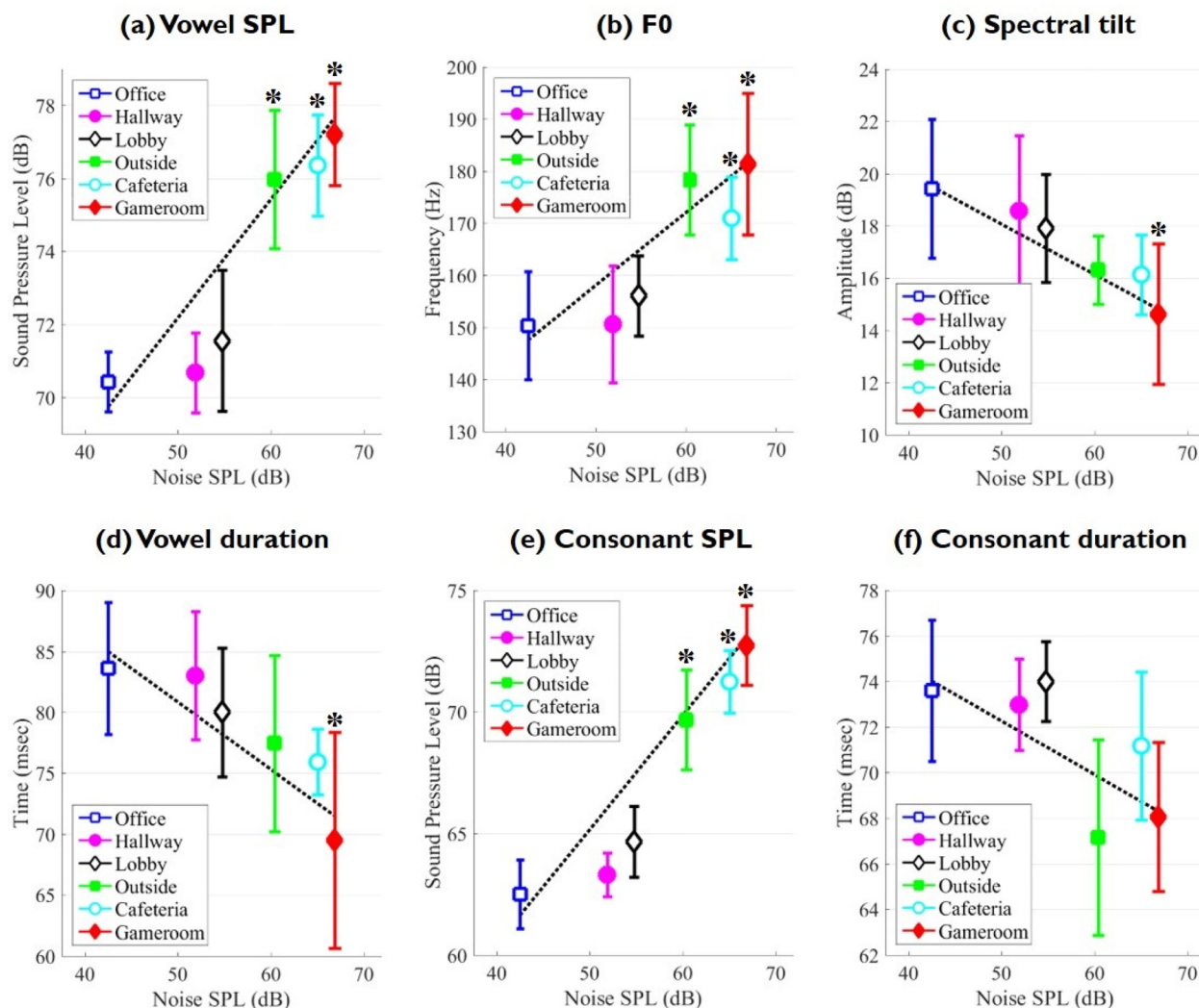

FIG. 5. (Color online) Acoustic analysis of vowel and consonant productions: individual variations of (a) vowel SPL, (b) fundamental frequency (F0), (c) spectral tilt, (d) vowel duration, (e) consonant SPL, and (f) consonant duration, as a function of varying noise SPL. Asterisk indicates statistical significance ($p < 0.05$) from neutral speech.

general, both features were altered by speakers under noisy environments. For consonant SPL, the results followed the similar pattern to vowel SPL. Two distinct groups were found, the high value (outside, cafeteria, and gameroom) and low value group (hallway and lobby). The high value group increased significantly from the office baseline ($p < 0.05$), while the low value group resulted in little or no change across both environments ($p > 0.05$). No significant pairwise differences were found between the conditions which belonged to the same group. For consonant duration, mean values monotonically decreased for most noisy conditions. However, there were no statistically significant difference in mean between quiet and all noisy conditions ($p > 0.05$). Speech produced in all noise environments had slightly shorter consonant duration than that of the baseline office environment.

2. Vowel-consonant ratios

Additional analyses were conducted on global shifts in acoustic features between individual phoneme classes. It has been previously demonstrated with NH listeners that a talker could maintain overall intensity, yet emphasize consonant phoneme class with respect to vowel class (House *et al.*, 1965; Hansen, 1988). Hansen (1988, 1996) also suggested that consonant duration increases at the expense of vowel duration in an effort to increase speech intelligibility under noise. In the present analysis, two different ratios were considered: (i) a consonant versus vowel intensity ratio (CVIR) and (ii) a consonant versus vowel duration ratio (CVDR) (House *et al.*, 1965; Hansen, 1996). Vowel and consonant intensities were computed using PRAAT software (Boersma, 2002), and phoneme duration was obtained from analysis of the phoneme-level transcripts. These ratios indicate how energy or duration between vowel and consonant speech classes changes under noisy conditions.

A pictorial representation of global shifts between individual phoneme classes is presented in Figs. 6(a) and 6(b). For each figure, shaded regions within each bar graph indicate average SPL/duration values for vowel and consonant phoneme classes. The asterisk indicates statistically significant shifts in mean based on the measures of CVIR and CVDR. Consider the CVIR first, where increased CVIRs resulted for most noisy conditions. The outside, cafeteria and gameroom conditions were significantly different from the office environment ($p < 0.05$). These particular changes in CVIR demonstrate increased consonant intensity as compared to vowel intensity in noisy condition. No significant differences were found for hallway and lobby as compared to office location ($p > 0.05$). For CVDR, no statistically significant shift in duration between vowel and consonant phoneme classes was observed in any locations ($p > 0.05$). It should be noted that consonant duration with respect to vowel duration plays a crucial factor in listener's ability to perceive the speech in the presence of noise (Hansen, 1988, 1996).

3. Vocal tract characteristics

Thus far, the investigation of the Lombard effect was focused on analysis of source excitation features. Speech SPLs, pitch, spectral tilt and duration are all controlled in some manner by the supra-glottal and glottal systems. It is reasonable to hypothesize that noise also affects the articulators that configure vocal tract shape. In order to investigate vocal tract shape, we considered only the vowel space, since it controls the analysis to fixed articulator positioning versus the more complex time varying requirements for liquids, glides, and diphthongs. For statistical reliability, phonemes with sufficiently large sample sets were considered. Four cardinal vowel nuclei, /a/, /æ/, /i/, and /u/, were chosen for this analysis. Phoneme-level transcription labels identified
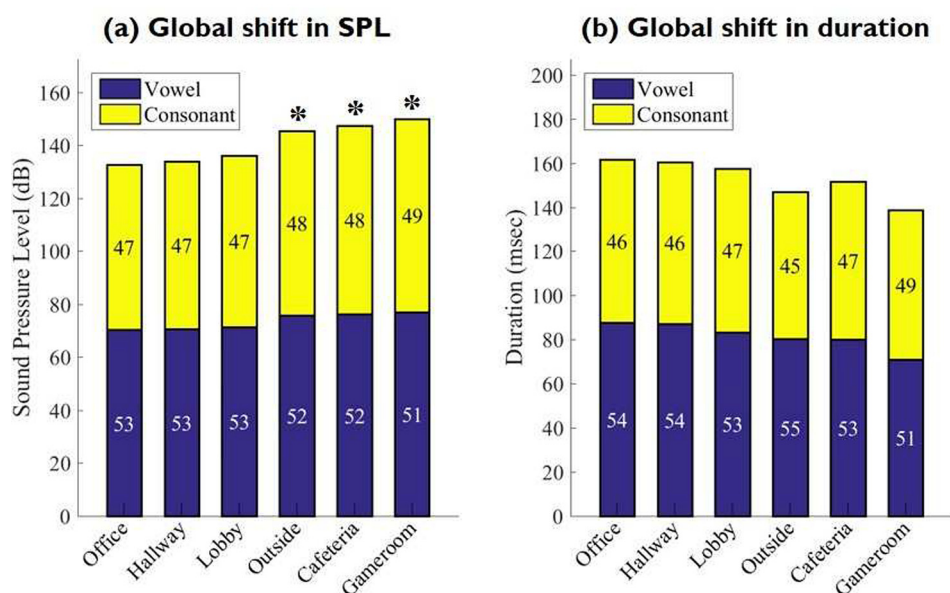

FIG. 6. (Color online) Pictorial representations of global shift in (a) SPL and (b) duration between vowel and consonant phoneme class. The speech class percentage is shown for each environment. Asterisks indicate significant shift in intensity/duration based on phoneme ratios.

J. Acoust. Soc. Am. **141** (4), April 2017

Lee *et al.* 2795

CHANGES IN PHONEME SPECTRAL STRUCTURE
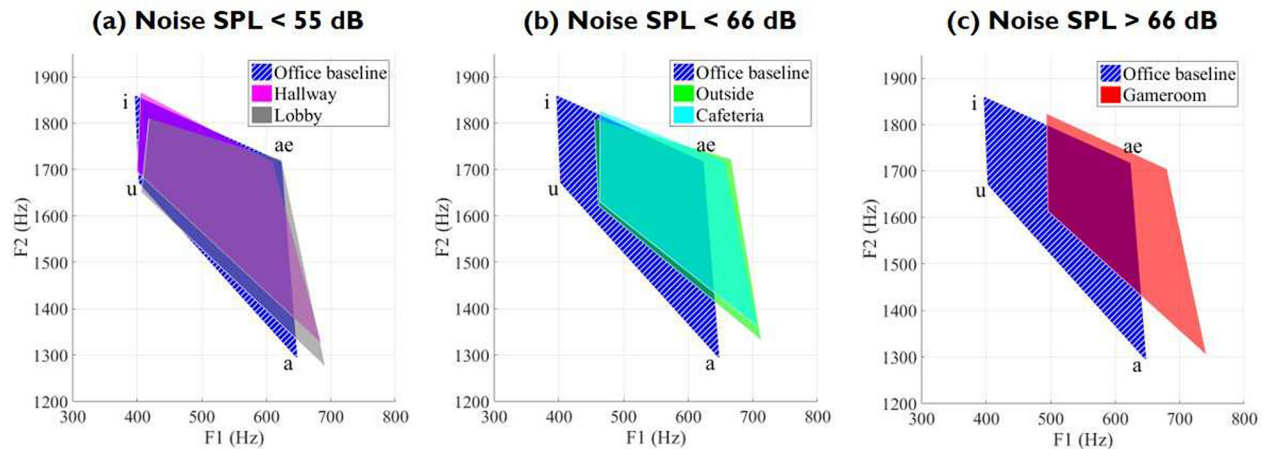## DUE TO ENVIRONMENTAL NOISE

FIG. 7. (Color online) Spectral characteristics of vocal-tract: plots of first formant frequency F1 versus second formant frequency F2 for vowel phonemes /a/, /æ/, /i/, and /u/ with respect to (a) hallway and lobby, (b) outside and cafeteria, and (c) gameroom environments. Office result was given in each plot for comparison.

boundary information of each phoneme. On average, more than 3000 instances of each phoneme were employed.

Figure 7 illustrates the vowel space for phoneme nuclei, /a/, /æ/, /i/, and /u/, under various noisy conditions. Figure 7(a) includes hallway and lobby, while Fig. 7(b) contains outside and cafeteria. Gameroom condition is shown in Fig. 7(c). Office baseline results are also provided in each figure for comparison. The abscissa in each figure denotes F1 formant locations, while the ordinate denotes F2 formant locations. Overall, F1 formant locations changed with respect to environment, but F2 formant location did not. Consider the front vowel /i/ first. F1 formant frequencies for /i/ phoneme significantly increased for outside, cafeteria and gameroom speech ($p < 0.05$). The pattern of the results was consistent with many other acoustic features (e.g., vowel/consonant SPLs and F0). However, this was not the case for the F2 feature. For the phoneme /i/, no significant changes occurred for F2 formant locations across all environments for all speakers ($p > 0.05$).

Similar changes were observed for the remaining three vowel phonemes, /a/, /æ/, and /u/. In these phonemes, shifts in formant locations generally followed the trend observed for the /i/ phoneme. For example, significant increases in F1 formant location for /u/ were found for environments: outside, cafeteria and gameroom conditions versus baseline office condition ($p < 0.05$). F1 formant location for /a/ and /æ/ phonemes were changed significantly for gameroom condition ($p < 0.05$). However, similar to the /i/ phoneme, F2 formant location also resulted in relatively small changes for /a/, /æ/, and /u/ phonemes across all environments ($p > 0.05$). Thus, F2 locations were not a major factor for changes in speech production in noisy environments. It should be noted that F2 formant frequency plays a critical role in speech comprehension of NH as well as CI users in noisy environments (Loizou, 2013).

C. Summary of Lombard speech features

In this study, analysis of speech production in six naturalistic environments was presented along with the Lombard

effect observed in speech produced by CI users. Statistical analysis techniques were employed to determine if any changes in speech features are reliable Lombard relayers. The results indicate that many speech features were used by the CI users in demonstrating Lombard effect stress condition. However, due to limited dataset and environments, it may be difficult to identify which specific features are sufficiently sensitive, and statistically reliable indicators of the Lombard effect. In order to identify these features, we grouped similar listening conditions together. Here, the environmental conditions were grouped into two areas: low and high noise groups. Low noise group includes office, hallway, and lobby, and high noise group includes outside, cafeteria, and gameroom. The decision was made based on average noise SPL [Fig. 3(a)] and average speech SPLs [Figs. 5(a) and 5(e)]. Next, ANOVA was performed on the two groups. Therefore, this section allowed for a manageable summary of important Lombard relaying features from naturalistic audio recordings by fixing two conditions.

The analysis indicates that speech SPLs and fundamental frequency (F0) were the primary stress relayers for varying vocal effort. Vowel ($F[1, 40] = 125.76, p < 0.001$) and consonant ($F[1, 40] = 116.43, p < 0.001$) classes resulted in an increased SPL as vocal effort increased. In addition to individual phoneme classes, the phonetic class intensity ratio (CVIR) identified a strong shift in energy from vowels toward consonants in Lombard speech ($F[1, 40] = 52.63, p < 0.001$). Thus, CVIR can be used as a reliable indicator for speech under varying noise. The average F0 value was found to be statistically different from neutral styles ($F[1, 40] = 50.00, p < 0.001$). Spectral tilt was also found to be a reliable perturbation of Lombard speech ($F[1, 40] = 15.37, p < 0.001$). Vowel ($F[1, 40] = 14.16, p < 0.001$) and consonant ($F[1, 40] = 9.00, p < 0.005$) durations were found to be valid discriminating features for Lombard speech as well. However, duration ratio between vowel and consonant (CVDR) could not be considered as a discriminating

Lombard indicator ($F[1, 40] = 1.99, p > 0.05$) at least on the environments tested. CVDR indicated only a slight shift in vowel duration towards an increase in consonant duration which was not statistically significant. Average first formant location F1 for the selected vowels /a/, /æ/, /i/, and /u/ showed statistically significant shifts from neutral speech ($F[1, 40] > 12.36, p < 0.001$). F2 formant location for /a/, and /u/ phonemes were significantly different from neutral speech as well ($F[1, 40] > 4.33, p < 0.05$). However, other phonetic areas such as second formant locations F2 for /i/, and /æ/ phoneme were not significantly different ($F[1, 40] < 1.42, p > 0.05$).

IV. DISCUSSION

A. The Lombard effect in CI users

Many speech production features identified in this study were observed to change in the presence of background noise. It was found that postlingually deafened CI users modify both segmental and suprasegmental properties of their speech in different listening environments. These features serve to calibrate speech production (i.e., the speaker monitors their relations between his/her own phonemic intention and their acoustic output in the presence of noise). Moreover, it also influences speech production along an instantaneous basis, thus speakers modulate at least some suprasegmental features of their ongoing speech gesture. The results from this study indicate that CI users exhibit the Lombard effect in adverse listening environments.

Many investigators have suggested that auditory information for NH listeners may be used to modulate at least some suprasegmental features of speech production under noise (Pisoni *et al.*, 1985; Hansen, 1988; Summers *et al.*, 1988; Junqua, 1992; Hansen, 1996; Lu and Cooke, 2008; Garnier *et al.*, 2010). In these studies, results showed an increase in overall amplitude of vocalic sections, increased duration, increased average F0, and a decreased spectral tilt. This was presumed to be the result of increased subglottal pressure, and vocal-fold tension as a response to the reduced auditory feedback due to noise. Existing studies suggest that the most widely considered area of the Lombard effect involves vocal intensity and F0 (Hansen, 1996; Lu and Cooke, 2009; Garnier *et al.*, 2010). Spectral balance of vowels was affected by the higher vocal effort for Lombard speech, resulting in relatively greater intensity in the higher frequency bands of the spectrum (Hansen, 1988; Junqua, 1992; Lu and Cooke, 2008).

Interestingly, the results showed that vowel and consonant duration [Figs. 5(d) and 5(f)] decreased in most noisy environments, in contrast to the earlier studies. It is well established in NH study that Lombard speech has a generally increased phoneme duration in comparison to normal speech, and speech intelligibility in noise is associated with lengthening the phoneme duration (Junqua, 1996; Lu and Cooke, 2008; Garnier *et al.*, 2010). We suggest that the main contribution to this difference is the conversational speaking style which was used in the data analysis. Participants in this study produced conversational speech in realistic scenarios, while the previous studies primarily focused on reading speech style

with given sentences. Another possibility is that increasing phoneme duration is not necessary for maintaining high intelligibility. Several studies confirmed that other inherent temporal properties, such as temporal amplitude modulations and vowel-consonant duration ratios, may directly contribute to enhanced intelligibility rather than phoneme durations (Payton *et al.*, 1994; Hansen, 1996; Krause and Braida, 2004).

In addition to suprasegmental variables, there has been a general consensus concerning the control of segmental features (Pisoni *et al.*, 1985; Hansen, 1988; Summers *et al.*, 1988). The rise in subglottal pressure needed to increase vocal effort leads to an increase in formant locations. For example, the wider jaw opening in order to increase sound amplitude causes an increase in F1 frequency (Huber and Chandrasekaran, 2006). It has also been suggested that under noisy conditions, speakers vary their speech characteristics so that speech segments rich in information are emphasized, while those less important to intelligibility are deemphasized (House *et al.*, 1965; Hansen, 1988; Sodersten *et al.*, 2005). For example, consonant energies increased at the expanse of vowel energy under noisy conditions in an effort to increase speech intelligibility (House *et al.*, 1965; Hansen, 1988). This is a useful characteristic, as consonants carry more speech information in the presence of noise.

The consistency between the two speaker groups (CI versus NH) indicated above could be mainly due to the presence of auditory feedback provided by the CI device. Long-term absence of auditory feedback could potentially result in poor regulation of acoustic, phonetic features of adventitiously deafened adults, such as F0, intensity, duration, etc. (Leder *et al.*, 1987; Lane and Webster, 1991). CI users, however, may demonstrate useful Lombard perturbation for regulating speech production features in noisy environments, which thereby assists in the development of more nearly neutral/typical acoustic, phonetic and temporal patterns under noise (Hochmair-Desoyer *et al.*, 1981; Kirk and Edgerton, 1983; Svirsky and Tobey, 1991; Svirsky *et al.*, 1992; Perkell, 2012).

The modification of speech production features under the Lombard effect may contribute to ensure intelligible communication in adverse noisy environments. The data from this study indicates that CI users respond to varying background noise types, and change their speech production accordingly. This articulatory modification allows speakers to avoid speech masked by the acoustic noise to compensate for the decreased SNR. Previous studies have reported that Lombard effect speech is more intelligible than speech under normal conditions (Summers *et al.*, 1988; Junqua, 1996; Lu and Cooke, 2008; Garnier *et al.*, 2010). In these studies, the intelligibility gain increased with increased in vocal effort (e.g., intensity, fundamental frequency, and spectral tilt) The data presented in this study suggest a potential perceptual benefit of the Lombard effect for CI users.

B. Future direction

The present study focused on the speech production of CI users in varying environment types. However, it does not address the nature and the extent of the Lombard effect as

J. Acoust. Soc. Am. **141** (4), April 2017

Lee *et al.* 2797

compared to the NH listeners. As a part of our future work, we suggest repeating the same data collection with NH individuals in the same environments to establish a one-to-one comparison of CI-NH pair and NH-NH pair. Moreover, while the present study focused on the role of auditory feedback provided in the context of CI systems, there has been no study of CI signal processing features that may play a role in auditory feedback. We feel further discussion on other CI sound processing factors, such as automatic gain control (AGC), adaptive dynamic range optimization (ADRO), frequency band sampling or virtual channels, represent a wider range of issues, which are beyond the current scope of this study. In theory, a supplementary study in future could explore these various factors within the context of the Lombard effect.

Furthermore, specific variations in speech production features due to the Lombard effect investigated here can be used to formulate new algorithms for improved intelligibility in noisy conditions. For example, strategies that exploit the impact of particular acoustic features of speech with respect to Lombard speaking style (Zorila *et al.*, 2012; Godoy and Stylianou, 2013). Historically, it is known that different environments will have specific noise types and levels. Traditional front-end processing for hearing aids and CIs, for example, have focused on noise suppression to minimize the impact of noise. Algorithmic advancements which modify neutral speech based on Lombard effect properties offers a unique opportunity to improve the listening/decoding experiences of CI users.

V. CONCLUSIONS

In this study, we analyzed the speech production of CI users with respect to environmental context. Naturalistic human-to-human voice interactions were captured using mobile personal audio recordings from continuous single-session audio streams collected over various realistic environments. An analysis of speech produced in noise and the Lombard effect observed in the speech of CI users was presented. The results indicated that speakers demonstrated increased vocal effort, including F0 and speech SPL, as well as altered glottal spectral slope, and phoneme duration in response to challenging noisy environments. Segmental articulatory movements, for example, F1 for specific phonemes such as /a/, /æ/, /i/, and /u/, also appeared to play an important role in relaying Lombard perturbation for speech produced in the presence of noise. The significance of the results is that the Lombard effect could potentially be helping CI users to ensure/maintain intelligible communication by compensating for the reduced SNR. The specific variations due to the Lombard effect can be leveraged for new algorithm development and further applications of speech technology to benefit CI users.

ACKNOWLEDGMENTS

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

Bořil, H., and Hansen, J. H. (**2010**). "Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments," IEEE Trans. Audio Speech Lang. Process. **18**, 1379–1393.

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. Audio Speech Lang. Process. **8**, 429–442.

Dorman, M., and Spahr, A. (**2006**). *Speech Perception by Adults with Multichannel Cochlear Implants in Cochlear Implants*, 2nd ed. (Thieme Medical Publishers, New York, NY), pp. 193–204.

Garnier, 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.

Gilkerson, J., and Richards, J. A. (**2008**). "The LENA foundation natural language study," technical report, LENA Research Foundation, Boulder, CO.

Godoy, E., and Stylianou, Y. (**2013**). "Increasing speech intelligibility via spectral shaping with frequency warping and dynamic range compression plus transient enhancement," in *Proceedings of the Annual Conference of the International Speech Communication Association*, 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. (**1996**). "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," Speech Commun. **20**, 151–173.

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. Process. **17**, 366–378.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Fischer, R. E. (**1981**). "Four years of experience with cochlear prostheses," Med. Prog. Technol. **8**, 107–119.

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.

Huber, J. E., and Chandrasekaran, B. (**2006**). "Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults," J. Speech Lang. Hear. Res. **49**, 1368–1379.

Iseli, M., Shue, Y.-L., and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am. **121**, 2283–2295.

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

Junqua, J.-C. (**1996**). "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex," Speech Commun. **20**, 13–22.

Kirk, K., and Edgerton, B. (**1983**). "The effects of cochlear implant use on voice parameters," Otolaryngol. Clin. North Am. **16**, 281–292.

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.

Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2007**). "Effects of short-and long-term changes in auditory feedback on vowel and sibilant contrasts," J. Speech Lang. Hear. Res. **50**, 913–927.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Lang. Hear. Res. **14**, 677–709.

Lane, H., and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**, 859–866.

Leder, S. B., Spitzer, J. B., and Kirchner, J. C. (**1987**). "Speaking fundamental frequency of postlingually profoundly deaf adult men," Ann. Otol. Rhinol. Laryngol. **96**, 322–324.

LENA Foundation (**2014**). "LENA research foundation," www.lenafoundation.org/ (Last viewed 9/12/2014).

Loizou, P. C. (**1999**). "Introduction to cochlear implants," IEEE Eng. Med. Biol. Mag. **18**, 32–42.

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

Lombard, E. (**1911**). "Le signe de l'elevation de la voix," Ann. Malad. 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**). "Speech production modifications produced in the presence of low-pass and high-pass filtered noise," J. Acoust. Soc. Am. **126**, 1495–1499.

Matthies, M. L., Svirsky, M., Perkell, J., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Lang. Hear. Res. **39**, 936–946.

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

Perkell, J. (**2012**). "Movement goals and feedback and feedforward control mechanisms in speech production," J. Neuroling. **25**, 382–407.

Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C., and Yuchtman, M. (**1985**). "Some acoustic-phonetic correlates of speech produced in noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tampa, FL, pp. 1581–1584.

Sangwan, A., Ziaei, A., and Hansen, J. H. L. (**2012**). "ProfLifeLog: Environmental analysis and keyword recognition for naturalistic daily audio streams," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, pp. 4941–4944.

Sodersten, M., Ternstrom, S., and Bohman, M. (**2005**). "Loud speech in realistic environmental noise: Phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers," J. Voice **19**, 29–46.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.

Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., and Miyamoto, R. T. (**2000**). "Language development in profoundly deaf children with cochlear implants," Psychol. Sci. **11**, 153–158.

Svirsky, M. A., and Tobey, E. A. (**1991**). "Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users," J. Acoust. Soc. Am. **89**, 2895–2904.

Vick, J. C., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., and Zandipour, M. (**2001**). "Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing," J. Speech Lang. Hear. Res. **44**, 1257–1267.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature **352**, 236–238.

Yu, C., Hansen, J. H. L., and Oard, D. W. (**2014**). " 'Houston, we have a solution': A case study of the analysis of astronaut speech during NASA Apollo 11 for long-term speaker modeling," in *Proceedings of the Annual Conference of the International Speech Communication Association*, Singapore, pp. 945–948.

Yuan, J., and Liberman, M. (**2008**). "Speaker identification on the SCOTUS corpus," in *In Proceedings of Acoustics*, Paris, France, pp. 5687–5690.

Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., and Feng, H. (**2008**). "Cochlear implants: System design, integration, and evaluation," IEEE Rev. Biomed. Eng. **1**, 115–142.

Ziaei, A., Sangwan, A., and Hansen, J. H. L. (**2012**). "Prof-Life-Log: Audio environment detection for naturalistic audio streams," in *Proceedings of the Annual Conference of the International Speech Communication Association*, Singapore, pp. 2514–2517.

Ziaei, A., Sangwan, A., and Hansen, J. H. L. (**2013**). "Prof-Life-Log: Personal interaction analysis for naturalistic audio streams," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, Canada, pp. 7770–7774.

Zorila, T.-C., Kandia, V., and Stylianou, Y. (**2012**). "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Proceedings of the Annual Conference of the International Speech Communication Association*, Portland, OR, pp. 634–637.