Published in final edited form as:

Lang Speech. 2018 June ; 61(2): 227–245. doi:10.1177/0023830917713775.

Supra-Segmental Changes in Speech Production as a Result of Spectral Feedback Degradation: Comparison with Lombard Speech

Elizabeth D. Casserly,

Trinity College, USA

Yeling Wang,

Trinity College, USA

Nicholas Celestin,

Trinity College, USA

Lily Talesnick, and

Trinity College, USA

David B. Pisoni

Indiana University, USA

Abstract

Perturbations to acoustic speech feedback have been typically localized to specific phonetic characteristics, for example, fundamental frequency (F0) or the first two formants (F1/F2), or affect all aspects of the speech signal equally, for example, via the addition of background noise. This paper examines the consequences of a more selective global perturbation: real-time cochlear implant (CI) simulation of acoustic speech feedback. Specifically, we examine the potential similarity between speakers' response to noise vocoding and the characteristics of Lombard speech. An acoustic analysis of supra-segmental characteristics in speaking rate, F0 production, and voice amplitude revealed changes that paralleled the Lombard effect in some domains but not others. Two studies of speech intelligibility complemented the acoustic analysis, finding that intelligibility significantly decreased as a result of CI simulation of speaker feedback. Together, the results point to differences in speakers' responses to these two superficially similar feedback manipulations. In both cases we see a complex, multi-faceted behavior on the part of talkers. We argue that more instances of global perturbation and broader response assessment are needed to determine whether such complexity is present in other feedback manipulations or if it represents a relatively rare exception to the typical compensatory feedback response.

Keywords

Speech motor planning; acoustic feedback; cochlear implant; Lombard speech; intelligibility

1 Introduction

Speech feedback perturbation studies have revealed a great deal about how speakers monitor and control their productions of segmental and non-segmental linguistic elements in real-time (e.g., Burnett, Freedland, Larson, & Hain, 1998; Houde & Jordan, 1998, 2002). However, the scope of perturbations in this research has been limited in terms of the focus on specific features: the majority of manipulations have targeted single, relatively isolated acoustic cues (e.g., fundamental frequency (F0), formant (F1), or formant (F2)). In this paper, we report the effects of a relatively novel global signal transformation—real-time simulation of cochlear implant (CI) processing—and compare it to those observed in the most well-known global feedback shift, Lombard speech.

The effect of CIs and their acoustic processing methods has been well-studied in the realm of speech perception. CIs use a relatively small number of tonotopically-organized electrodes to transmit acoustic signals directly to the auditory nervous system, resulting in a loss of fine-grained spectral detail relative to normal transduction (Niparko, 2009). This transformation of acoustic signals has been simulated (or “vocoded”) for normal-hearing listeners, causing them to mirror the difficulties experienced by clinical CI users in speech perception (Fu, Nogaki, & Galvin, 2005; Fu, Shannon, & Wang, 1998; Loebach, Pisoni, & Svirsky, 2009; Qin & Oxenham, 2003; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), environmental sound recognition (Shafiro, 2008; Shafiro, Sheft, Gygi, & Ho, 2012), and use of phonological working memory (Burkholder, Pisoni, & Svirsky, 2005; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000), among other areas. Simulation of CI processing has been used in these studies to isolate signal-related limitation in recognition (e.g., Shannon et al., 1995) and to test the potential efficacy of interventions designed to improve recognition accuracy (e.g., Loebach et al., 2009). Recently, however, a device was developed that was capable of performing the CI-simulation transformation in real-time, via a lightweight portable interface (Casserly, Pisoni, Smalt, & Talavage, 2011). This device made it possible to apply noise-modulated CI simulation to speech *production*, as well as perception.

1.1 Lombard speech

In the Lombard phenomenon, acoustic speech feedback is manipulated via either the addition of noise or the reduction of feedback amplitude/sidetone (Lane & Tranel, 1971). These perturbations result in highly systematic and well-studied changes to speech production collectively known as the Lombard reflex or sign: speakers increase speech amplitude, reduce speaking rate, hyperarticulate segmental acoustic contrast, and produce average fundamental frequencies outside their usual ranges (Garnier, Henrich, & Dubois, 2010; Lane & Tranel, 1971; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). These responses are highly consistent across talkers and seem to be largely independent of conscious control (Lane & Tranel, 1971; Pick, Siegel, Fox, Garber, & Kearney, 1989).

Of the changes seen in Lombard speech, the increase in amplitude is the most well-described and commonly known (see, e.g., Beckford Wassink, Wright, & Franklin, 2007). It links directly with the initial feedback amplitude manipulation and offers an intuitive explanation for the behavioral response: speakers are compensating for a mismatch between the

amplitude feedback they expect and the signal they actually receive. Indeed, the degree of noise masking or sidetone reduction correlates to a large degree with the magnitude of changes in speakers' voice amplitude (Lane & Tranel, 1971). The correlation is not perfect, however, and there is reason to believe that the Lombard effect is more complicated than a compensation-based account implies (Lane & Tranel, 1971; cf. Vogel, Fletcher, & Maruff, 2014).

For instance, the other domains of change seen in Lombard speech are less clearly connected to signal factors or feedback-error compensation. Speaking rate and F0 are relatively preserved when noise is added to speech feedback, and vowel formants would be evenly masked across the F1/F2 range (for broadband noise manipulations, at least). Alterations to these characteristics therefore do not lessen feedback error. Instead, they may represent speakers' attempt to make their perceptual goals easier to monitor in other ways. That is, speech produced with these characteristics, and under Lombard conditions in general, is more intelligible than speech produced under typical conditions (Cooke, Mayo, & Villegas, 2014; Garnier et al., 2010; Patel & Schell, 2008; Summers et al., 1988). In order to control their speech motor system effectively, speakers need to be able to perceive the accuracy of their productions overall, not just to detect and minimize deviations from expected sensory feedback. Producing highly intelligible speech helps speakers maintain accuracy even under challenging noise-filled listening conditions, and that may be the point of the nonamplitude alterations seen in Lombard speech.

1.2 Speech with CI-simulated feedback

The initial prediction for speech produced with CI-simulation used as an acoustic feedback perturbation was that it would show error-related compensations similar to those observed in other, less complex feedback perturbations (e.g., shifting of F0 or vowel formants; Elman, 1981; Houde & Jordan, 1998). Specifically, speakers were expected to compensate for changes in their spectral structure. Casserly (2015) tested this prediction, examining domains of segmental acoustic-phonetics that rely on production of precise spectral frequencies: sibilant fricative centroids and vocalic F1/F2 qualities. Neither area showed the predicted compensation effects for simulationbased shifting of average frequency: fricatives were not altered with CI-simulated feedback at all, and vowel quality alterations bore no relationship to the specifics of the spectral manipulation with respect to each speaker's particular F1/F2 vowel space (Casserly, 2015). Instead of compensation for relative frequency shifts, speakers significantly reduced phonological contrast in F1 while maintaining consistent F2 production. Individual differences in the means of reducing F1 contrast were observed—some speakers selectively lowered high vowels while others raised low vowels—but these patterns were not predictable based on segmental acoustic characteristics. Overall, therefore, it appeared that speakers did not respond to the CI-simulated feedback by compensating for its specific signal manipulation effects.

Given that CI processing was simulated using amplitude-modulated bands of white noise, however, and it generally results in a wide-spread degradation of spectral structure, CI simulation may bear more similarity as a feedback perturbation to the instruction of noise in Lombard speech than to localized shifting of F0 or vowel formants.

1.3 Current research

In this paper, we report the results of three investigations into the supra-segmental effects of CI-simulation as an acoustic feedback perturbation. First, we analyze sentence-level speech for shifts in speaking rate, F0 production, and voice amplitude—domains that have not been investigated in the response to CI-simulated feedback and which are classic areas of change in the Lombard sign. Then we report the results of two perceptual studies that directly assess the intelligibility of speech produced with CI-simulated feedback. Together, these data enable us to evaluate the similarity of the Lombard effect and the effect of real-time CI simulation in terms of their broader, non-compensatory impact on speech production.

2 Analysis of speech corpus in Casserly (2015)

The reduction of F1 contrast seen in Casserly (2015) with CI-simulated speech feedback runs counter to what would be expected in Lombard speech, which is typically characterized by expansion of vocalic contrast (Cooke et al., 2014; Summers et al., 1988). However, Lombard speech is more than just vowel hyperarticulation; as discussed above, it involves changes to supra-segmental speech attributes that have not been investigated previously. Moreover, there is reason to believe that these attributes might be good candidates for Lombard-like alterations in speech with CI-simulated feedback. That is, CI-simulation is a manipulation of multiple, quasi-independent acoustic cues. Spectral resolution is greatly reduced by the reduction in channels of information, but timing and amplitude cues are preserved (cf. Shannon et al., 1995). Also, the spectral degradation is not uniform across the frequencies associated with formants versus pitch: simulations (and real CIs; Niparko, 2009) transmit only acoustic signals falling between approximately 200 and 7000 Hz, eliminating typical F0 ranges. Frequencies outside this band, however, can still be transmitted to the cochlea via bone conduction in normal-hearing participants (Reinfeldt, Ostli, Håkansson, & Stenfelt, 2010; Stenfelt & Håkansson, 2002). Speakers experiencing CI-simulated feedback, therefore, receive highly degraded mid-frequency information that masks bone-conducted signal in these ranges, but they also experience unmasked, competition-free, bone-conducted F0 cues.

Despite the contrast reduction seen in vowel acoustic-phonetics, therefore, it is possible that speakers produced Lombard-like changes in speech domains where perceptual cues were most reliable: rate; F0; and amplitude. Below, we test this possibility directly by re-examining the Casserly (2015) corpus, specifically sentence-level speech that has not been analyzed previously. Speech was collected from participants under three conditions: with normal acoustic feedback (Baseline); immediately after introducing a real-time CI-simulator as an acoustic feedback manipulation (Epoch 2); and following 15 minutes of naturalistic experience with CI-simulated acoustics in a one-on-one conversation with a researcher (Epoch 3). The segmental acoustic-phonetics of isolated words produced across those conditions were reported in Casserly (2015), as well as the segmental phonetics from a small group of control talkers' speech.

For the present analysis, sentence materials enabled us to effectively target four supra-segmental acoustic cues: speaking rate; average F0; F0 variability; and voice amplitude. If speakers produced speech with Lombard-like alterations, we would expect to see reduced

speaking rate, also associated with clear speech modes (Hazan & Baker, 2011; Vogel et al., 2014), an increase in average F0, greater variation in F0, and greater amplitude (Cooke et al., 2014; Garnier et al., 2010; Lane & Tranel, 1971; Summers et al., 1988).

2.1 Design and methods

2.1.1 Subjects.—The speech of twelve adults (two males, and ten females) with no history of speech or hearing disorders was examined in this study. All subjects were monolingual native speakers of American English and were between 18 and 30 years of age (mean age 21.1 years). Nine subjects completed the experimental protocol (including the two males), while the remaining three (all female) served as controls. The study protocol was reviewed and approved by the Institutional Review Board at Indiana University, and all subjects gave their informed consent to participate.

2.1.2 Acoustic feedback transformation.—Simulation of cochlear implant processing was carried out using a portable, real-time vocoder (PRTV), a device consisting of a central processor (8 GB iPod touch, model A1367), input microphone (Williams MIC090 mini lapel clip), noise-isolating earphones with disposable single-use foam tips (Etymotic HF5), and a set of noise-attenuating headgear (Elvex SuperSonic) worn simultaneously with the insert earphones. Custom PRTV software took continual sound input from the lapel microphone, worn on the attenuators just above a subject's left ear, applied the acoustic transformation (cf. Kaiser & Svirsky, 1999), and relayed the modified acoustic signal to the insert earphones with less than 10 milliseconds (ms) delay (Casserly et al., 2011).

The signal processing transformation is summarized in Figure 1. Sampled acoustics were subjected to a broad band-pass frequency filter of 252–7000 Hz then split into eight non-overlapping frequency bands (Casserly, 2015; Kaiser & Svirsky, 1999). The amplitude within each of these channels was then used to generate envelope-matched noise of the same bandwidth. The eight amplitude-matched noise channels were then summed to create a noise-vocoded acoustic simulation of a CI. Resolution within each channel was lost during the transformation, therefore, along with any information occurring at frequencies higher than 7000 Hz or lower than 252 Hz, resulting in degradation of fine-grained spectral information in a reduced range.

2.1.3 Procedure—Speech samples were recorded in a sound-attenuating booth (Industrial Acoustics Company, Inc.) over three 20 minute sessions or epochs. During each recording epoch, speakers were seated in front of a computer monitor showing orthographic prompts and a desk with a stand microphone (Audio Technica AT2021). It was necessary to use a stand, rather than head-mounted, microphone for recording because the feedback-altering PRTV, worn by speakers in the feedbackalteration conditions, was also head-mounted and left no room for an additional microphone set up. The stand microphone accurately recorded speakers' F0 and speaking rates, but measures of speech amplitude may be conflated with changes in the physical distance between the speaker and the microphone across recording blocks.

Speakers were asked to read materials aloud while their speech was recorded and an experimenter live-monitored the session. Stimuli were the same across all three epochs, presented in random order with pre-programmed timing. All recordings were made at 44.1 kHz (16 bit) and stored digitally for analysis.

In the first, baseline recording epoch, no transformation of acoustic feedback occurred and subjects did not wear the PRTV. Immediately following the baseline epoch, experimental subjects were fitted with the PRTV. After the real-time acoustic transformation was begun, subjects were instructed to remain quiet until the recording session began. Speech produced in Epoch 2, therefore, represented the first sound subjects heard through the novel feedback transformation. After completing Epoch 2, subjects exited the recording booth and engaged in 15 minutes' naturalistic conversation with the experimenter. For experimental subjects, all acoustic signals continued to be transformed in real-time for subjects during this "break," including their own speech and that of the experimenter. Following this brief exposure period, subjects re-entered the recording booth and completed the final recording epoch, still wearing the PRTV.

Control subjects completed an identical protocol, including 15 minutes' conversation with the experimenter, but never experienced any modification of their acoustic feedback and did not wear the PRTV device. The control group therefore allowed us to address the potential effects of repeated production of our materials. That is, repetition of items is known to affect speech in terms of hyper- or hypo-articulation (Lindblom, 1990) and prosodic focus or emphasis (Baker & Bradlow, 2009), and materials from this experiment were produced across three repetitions by our speakers. We elicited these repetition-related changes directly with the control group, asking them to complete the same three-session protocol as our experimental talkers but without any manipulation of their acoustic feedback. By comparing supra-segmental cues between the two groups, we can differentiate genuine effects of feedback perturbation from those of mere repetition.

2.1.4 Stimulus materials.—All subjects read aloud 114 isolated English words and a set of 12 meaningful, "phoneme-specific" English sentences. Segmental acoustic phonetics were analyzed using stressed vowels in the isolated word data set (described in more detail in Casserly, 2015), but the supra-segmental characteristics of single-word stimuli are limited. We therefore examined rate, F0, and amplitude using the sentences. These were originally designed to heavily load their segmental content towards particular phonological classes; for example, *Nine men were moaning all morning* (Huggins & Nickerson, 1985). Speakers were given a total of four seconds to produce each sentence. Additional stimulus materials, including a short passage of connected text, an additional set of isolated words, and an additional set of meaningful sentences with prosodic constraints, were also read by all speakers but were not used for any of the analyses reported below or in Casserly (2015).

2.1.5 Design and analysis.—Four supra-segmental speech production properties were analyzed in the test sentence recordings: utterance duration (inversely related to speaking rate); mean F0; the standard deviation (*SD*) of F0 for each utterance, as a measure of F0 variability; and mean root mean square (RMS) amplitude. Variability can be assessed in multiple ways; we chose the *SD* statistic because of its straightforward and familiar

interpretation. All of the analyses concerning F0 variability below were also conducted with a coefficient of variation statistic instead of *SD* and the same results were obtained.

Utterance duration was coded as the duration of each utterance from the onset of the speech signal to its final offset. Statistical means and *SDs* of F0 were calculated from continuous estimates of F0 over the course of an utterance, generated by autocorrelated F0-tracking (range limits 75–300 Hz) in Praat (Boersma & Weenink, 2008). Samples within individual utterances with no estimate of F0 were not included in either calculation. Mean RMS amplitude was calculated from the onset of each target utterance to its conclusion without editing. Pauses, silences associated with voiceless stop closures, etc. were therefore included in utterance means.

Differences in production of these cues across recording epochs and groups were assessed via mixed-design repeated-measures analysis of variance (RM-ANOVA) with Group as a betweensubjects factor and Epoch as a within-subjects factor. This analysis allowed particular materials to be compared across epochs, controlling for variations in phonetic context (e.g., number of words affecting utterance duration or particular segment classes affecting automated F0-tracking).

2.2 Results

2.2.1 Speaking rate.—Mean sentence durations for speakers in the experimental and control groups are shown in Figure 2 for the baseline and recording Epochs 2, and 3. Subjects in the experimental group had average utterance durations of 1.648 seconds (s), 1.854 s, and 1.709 s, respectively. Subjects in the control group produced sentences of 1.705 s, 1.666 s, and 1.597 s. RM-ANOVA revealed a significant main effect of Epoch, $F(2, 140) = 8.025$, $p < 0.001$, $\eta^2 = 0.103$, and a significant Epoch $\times$ Group interaction, $F(2, 140) = 10.047$, $p < 0.001$, $\eta^2 = 0.126$; the effect of Group overall was not significant ($p > 0.05$, $\eta^2 = 0.011$). Post-hoc Bonferroni analyses examining the Epoch main effect showed that utterances produced in Epoch 2 were significantly longer than those from either Epoch 1 or 3 (p 's < 0.05), but this was most likely not consistent across members of both groups. To examine the Epoch $\times$ Group interaction, within-group RM-ANOVAs were conducted.

Repeated-measures analysis of variance for speakers in the experimental group returned a significant main effect of Epoch, Huynh–Feldt corrected $F(2, 90) = 43.901$, $p < 0.001$, $\eta^2 = 0.494$. Bonferroni-corrected pairwise comparisons showed that durations in all three Epochs were significantly different from one another (all p 's < 0.01), with an increase in utterance length between Epochs 1 and 2, when feedback perturbation was introduced, and a decrease in length to an intermediate value occurring between Epochs 2 and 3, as speakers gained more experience. RM-ANOVA for speakers in the control condition also returned a significant effect of Epoch, $F(2, 30) = 5.271$, $p = 0.011$, $\eta^2 = 0.260$, but in this case only the pairwise decrease in duration between Epochs 1 and 3 was significant (Bonferroni, $p < 0.05$; all other p 's > 0.05). The interaction between Epoch and Group, therefore, appears to stem from the different directions of change across Epochs, with speakers in the control condition gradually decreasing their utterance durations (increasing speaking rate), while speakers in the experimental condition significantly increased utterance duration (reducing speaking

rate) with the introduction of the feedback perturbation, followed by a partial return to baseline.

2.2.2 Average F0.—Mean F0 for all speakers in the experimental group was 189.6 Hz in Epoch 1, 193.7 Hz in Epoch 2, and 184.1 Hz in Epoch 3. Speakers in the control group (all female) had mean F0s of 196.1 Hz, 193.8 Hz, and 201.1 Hz, respectively. Both sets of means are shown in Figure 3. A RMANOVA conducted on the F0 data returned a significant Epoch $\times$ Group interaction, Huynh–Feldt corrected $F(2, 138) = 6.314$, $p = 0.004$, $\eta^2 = 0.084$, with no main effects of Epoch, Huynh–Feldt corrected $F(2, 138) = 0.128$, $p = 0.858$, $\eta^2 = 0.002$, or Group, $F(1, 69) = 0.675$, $p = 0.414$, $\eta^2 = 0.010$.

Given the relationship between speaker sex and F0, this analysis was also conducted on only the female speakers in both groups (experimental $n = 7$, control $n = 3$) in order to control for gender differences in the group composition. The female-only analysis returned the same Epoch $\times$ Group interaction, Huynh–Feldt corrected $F(2, 114) = 6.60$, $p = 0.003$, $\eta^2 = 0.104$, but also a significant effect of Group, $F(1, 57) = 5.092$, $p = 0.028$, $\eta^2 = 0.082$, with experimental talkers having a lower F0 across Epochs than controls.

Simple effects analysis of the control group speakers did not show statistically significant changes in mean F0 across epochs, $F(2, 32) = 2.110$, $p = 0.138$, $\eta^2 = 0.117$. An RM-ANOVA over all experimental speakers, male and female, did show significant changes across feedback conditions, Huynh–Feldt corrected $F(2, 106) = 7.818$, $p = 0.001$, $\eta^2 = 0.129$, and the same differences were seen in a female-only sub-analysis as well, Huynh–Feldt corrected $F(2, 70) = 12.033$, $p < 0.001$, $\eta^2 = 0.256$. Pairwise post-hoc Bonferroni comparisons showed that the differences between each Epoch were significant (all p 's < 0.05) such that there was an initial increase in F0 when CI-simulation was applied to speakers' feedback followed by a decrease to F0 levels significantly lower than those produced at baseline (Figure 3).

2.2.3 F0 variability.—Initial analysis of all speakers' F0 variability showed a significant effect of Group, $F(1, 69) = 4.839$, $p = 0.031$, $\eta^2 = 0.066$, but a non-significant Epoch effect, $F(2, 138) = 0.693$, $p = 0.502$, $\eta^2 = 0.010$, and a non-significant interaction, $F(2, 138) = 2.614$, $p = 0.077$, $\eta^2 = 0.037$. The Group difference revealed in this analysis showed that experimental speakers' variability was less than that of control speakers. Speakers with lower F0, however, tend to have smaller ranges as well, which in turn affects variability. Given that our experimental group contained not only the male speakers in the study but also a set of female speakers with significantly lower F0 than those in the control (Section 2.2.2), this Group difference was not unexpected. In order to minimize the possible effects of this confound on the data, therefore, a parallel RM-ANOVA was also conducted on a data set of only the female speakers in both groups.

Figure 4 shows the average variability in F0 for female subjects in the experimental and control groups across the three recording epochs. RM-ANOVA of the F0 data from female subjects in both groups showed that the Group differences were no longer significant, $F(1, 57) = 0.895$, $p = 0.348$, $\eta^2 = 0.015$, but that the Epoch $\times$ Group interaction was significant, $F(2, 114) = 3.659$, $p = 0.029$, $\eta^2 = 0.060$, albeit with a small effect size.

Simple effects RM-ANOVA on F0 variability in the full experimental speaker data set did show a significant main effect of Epoch, $F(2, 106) = 6.105$, $p = 0.003$, $\eta^2 = 0.102$, which was echoed in analysis of the female-only subset, $F(2, 70) = 9.884$, $p < 0.001$, $\eta^2 = 0.220$. The parallel within-group analysis of control speakers' F0 variability, by contrast, did not show significant change across Epochs, $F(2, 28) = 0.360$, $p = 0.701$, $\eta^2 = 0.025$. Pairwise comparison of F0 variability in experimental speakers showed there was a significant drop in variability between baseline ($SD = 36.1$ Hz) and Epoch 2 ($SD = 27.8$ Hz) when they initially encountered the CI-simulation feedback perturbation (see Figure 4).

2.2.4 Voice amplitude.—Mean RMS amplitudes across all sentences for all talkers were 66.1 dB at baseline, 66.9 dB in Epoch 2, and 65.2 dB in Epoch 3. Mixed-effects RM-ANOVA on these amplitude data returned a significant Group $\times$ Epoch interaction, Huynh–Feldt corrected $F(2, 121) = 6.64$, $p = 0.003$, $\eta^2 = 0.087$, but no significant differences overall across Epochs, Huynh–Feldt corrected $F(2, 121) = 2.33$, $p = 0.109$, $\eta^2 = 0.032$, or between Groups, $F(1, 70) = 0.136$, $p = 0.714$, $\eta^2 = 0.002$.

Simple effects analysis of changes across Epochs in each group showed that the differences with feedback perturbation were significant for speakers in the experimental group, $F(2, 90) = 58.3$, $p < 0.001$, $\eta^2 = 0.565$, but not for those in the control, Greenhouse–Geisser $F(2, 30) = 2.43$, $p = 0.125$, $\eta^2 = 0.140$. Mean utterance amplitudes for experimental speakers were 66.2 dB at baseline, 67.0 dB after the feedback perturbation was introduced, and 64.63 dB following the short learning period. Each of these amplitude levels was significantly different from the others (Bonferroni, p 's < 0.05). These results are summarized in Figure 5.

2.3 Discussion

Sentence-level speech from the Casserly (2015) corpus was analyzed in the supra-segmental domain in order to test the possibility that speakers showed Lombard-like effects in rate, intensity, and (bone-conducted) F0, which were all preserved under the feedback transformation. Significant changes under CI-simulation were observed in each of the four supra-segmental speech characteristics we assessed. Relative to their baseline speech production, subjects experiencing CI-simulation (1) reduced their speaking rate, (2) altered their average F0, initially raising and then lowering it significantly, (3) adopted a more monotone F0, and (4) altered their voice amplitude in parallel with mean F0, initially raising and then lowering relative to baseline. These alterations were distinct from simple repetition effects. That is, control speakers did not show any significant changes in average F0, F0 variability, or amplitude, and they altered their speaking rate to be faster, rather than slower, over the three repetitions of the sentence stimuli.

The supra-segmental CI-simulation responses parallel those seen in the Lombard effect in some ways, but differ substantially in others. Results in speaking rate provide the most straightforward correspondence: just as in Lombard speech (Cooke et al., 2014; Summers et al., 1988), speaking rate was reduced overall in speakers hearing noise-vocoded feedback. This is despite the fact that rate and relative timing are transmitted faithfully through the CI transformation. The rate changes cannot, therefore, be construed as a direct compensation

for any aspect of the feedback shift—just as the slowing seen in Lombard speech does not compensate for the addition of noise.

The situation is murkier for changes in mean F0 and amplitude. In both domains, speakers under CI-simulation initially produced shifts identical to those associated with Lombard speech by raising mean F0 and amplitude. Following the brief opportunity for perceptual learning, however, responses in both domains took on the opposite quality: participants spoke more quietly than at baseline and with a lower mean F0. The initial speech characteristics support the hypothesis that speakers are treating their feedback as similar to undifferentiated noise, but the change with experience implies that this response may be short-lived or stem from a different mechanism entirely.

Similar changes with experience were also observed in F0 variability, where speakers initially produced significantly less variation than at baseline but partially returned to baseline levels following perceptual learning. The lack of significant contrast between either baseline or initial CI-simulated epochs makes this final set of utterance variabilities difficult to interpret; once again, the clearest results were seen in the initial exposure period to the CI-simulation.

Unlike other measures, however, the reduction in F0 variability runs counter to the expectations for Lombard-like speech, where variation in F0 increases. Although deaf CI users typically cannot perceive F0 directly from their devices, the normal-hearing speakers in this study had access to bone-conducted feedback for F0. They therefore could have detected this monotone delivery style but did not immediately correct it to return to their baseline prosodic variation.

Taken together, the above re-analysis of speech collected in Casserly (2015) shows tentative support for the hypothesized links between CI-simulated feedback effects and Lombard speech in terms of the speaking rate, mean F0, and voice amplitude domains. The majority of similarities disappear after a short perceptual learning period for speakers, however, and the initial reduction in F0 variability is the opposite of what is typically seen in Lombard speech. We believe there are two possible interpretations for these results, therefore. First, speakers may have initially responded to CI-simulated feedback as though it were unstructured noise (i.e., with Lombard-like speech), but the brief period of perceptual learning was sufficient for them to detect the spectral structure and alter their response to something more complex. Or alternatively, speakers' initial response was entirely distinct from the Lombard reflex (as revealed by the shift to more monotone delivery) and the parallels in rate, amplitude, and mean F0 were due to chance (as shown by the reversals in two out of three domains in Epoch 3).

If the first interpretation is correct, then speech from the initial exposure period (Epoch 2) should at least parallel Lombard speech in another crucial dimension: increased intelligibility. If the response stems from a distinct mechanism, however, the changes in supra-segmental phonetics may not have benefits for intelligibility. The perceptual studies below test this difference in predictions.

3 Intelligibility of speech with CI-simulated feedback

Speech intelligibility is commonly assessed via a perceptual study where naïve participants are asked to complete an open-set speech recognition task. Differences in recognition accuracy or speed reflect differences in intelligibility. It can be difficult, however, to assess recognition accuracy with sentence-level materials; participants may use syntactic and semantic context to fill in words, for example, rather than recognizing them directly, and metrics must be specified for assessing linear order and morphological agreement in addition to lexical retrieval (see, e.g., keyword scoring in Loebach & Pisoni, 2008). The usefulness of reaction time data from sentence recognition is similarly limited, given the complexity of the response. Scoring and interpretation of the accuracy data is more straightforward with isolated word stimuli. The brevity of single words also enables each listener to respond to a greater number of tokens, spread among a wider set of talkers. Given the variability in perceptual abilities across listeners and the differences in intelligibility across talkers, such consistency is extremely important in producing interpretable results.

In order to assess the intelligibility of speech produced at Baseline and in Epoch 2, therefore, we returned to the set of isolated word productions analyzed acoustically in Casserly (2015). Acoustic analysis of the subset of tokens used for stimuli in this study (see below for details) showed that they contained the same supra-segmental changes observed in the sentence-level analysis in Section 2 above. For experimental talkers, mean word duration increased from 599 ms at Baseline to 678 ms in Epoch 2, mean F0 increased from 201 Hz to 204 Hz, F0 *SD* decreased from 23.8 Hz to 19.9 Hz, and voice amplitude increased from 64.8 dB to 65.6 dB, respectively. Control talkers' changes from Baseline to Epoch 2 were in the opposite direction for mean F0, F0 variability, and speaking rate. Given the consistency in supra-segmental acoustic change between these materials and sentence-level speech, we elected to use isolated words to test intelligibility, rather than sentences. We also asked participants to complete an explicit intelligibility discrimination task in addition to standard open-set recognition. In the discrimination task, two tokens of a single word were presented—one from Baseline and one from Epoch 2—and participants were asked to choose which token “was easier to understand.” This task assesses listeners' judgments about intelligibility, rather than intelligibility itself, but it is more sensitive to small differences in speech quality than a recognition task, and is informative even if participants are 100% accurate in their recognition performance. Both tasks were interpreted together in order to determine intelligibility differences across feedback conditions.

3.1 Method

3.1.1 Subjects.—Thirty-nine native speakers of English were recruited from Trinity College, in Connecticut, USA. Listeners were compensated for their participation via either a \$10 gift card to a local café, or course credit. Twenty-seven of the participants were female (69.2%), and 12 were male (30.7%). with ages ranging from 16 to 22 (mean 19.05 years). Participants were aware of the purpose of the study and gave full informed consent before participating. None of the participants reported a history of hearing problems, and all passed a hearing screening (calibrated Ambco 650A) with thresholds ≤ 30 dB HL between 500 and 8000 Hz before they participated in the experiment.

3.1.2 Design.—Listeners completed three blocks of trials. Blocks 1 and 3 were explicit discrimination trials, and Block 2 consisted of open-set recognition. Across all blocks, listeners heard speech from both experimental and control talkers; critical response patterns differentiated between these two talker types, that is, they were dependent on talkers' feedback contexts, and revealed a relative intelligibility benefit for speech in one condition over the other.

3.1.3 Stimulus materials.—A total of 96 word types were included, all rated as highly familiar, with balanced high and low frequencies-of-occurrence (195 and 5 tokens/million, respectively; Nusbaum, Pisoni, & Davis, 1984). All 96 words appeared in each block of trials, with listeners hearing tokens from both speaking contexts (Baseline and Epoch 2) back-to-back in random order in the discrimination task, and responding to tokens from both contexts in full randomization (192 tokens) in the recognition task. Talkers were represented equally; eight sets of two stimuli from each talker were used in each discrimination block and 16 tokens from each talker were used in recognition. The stimuli in the first discrimination task were identical to those in the following recognition task—only the trial procedure and response mode differed. New tokens of the 96 words, from different talker/word pairings, were introduced for the final discrimination block. The assignment of talkers to words in both sets of stimuli was random, with the conditions that no assignment was repeated in both sets and that each talker provided four high and four low frequency-of-occurrence words in each set.

3.1.4 Procedure.—Each participant completed the study in a quiet testing room, using a personal computer with a 19-inch monitor running E-Prime software. Stimuli were presented via headphones (Sennheiser HD 280 Pro) at an average sound pressure level of 70 dB. In discrimination trials, participants were told to indicate which of two tokens was “easier to understand.” The target word was displayed orthographically on the screen during presentation of both tokens, with an inter-stimulus interval of 750 ms. Participants pressed “f” to indicate a preference for the first token and “j” for the second. Presentation order of the two tokens was randomized on each trial for each subject. In recognition trials, an empty text field was provided for listeners' transcription of each token. Short breaks were provided between each testing block, and overall the study took participants less than an hour to complete.

3.1.5 Data analysis.—Participants' responses were recorded for all trials. For discrimination trials, the rate at which listeners selected tokens from each speaking context was calculated for each block of trials and for each talker condition (experimental vs. control); recognition was scored for accuracy, with correct responses and obvious orthographic errors (e.g., “manuvor” for target “maneuver”) receiving full credit and all other responses marked as incorrect. Discrimination data were analyzed using one-sample *t*-tests to determine whether preference rates in each data set differed significantly from chance (50%), and an RM-ANOVA was used to assess differences in preference rates across talker condition and test block. Recognition accuracy was analyzed via a two-factor RM-ANOVA looking for differences across speaking contexts (baseline vs. perturbation) and across talker conditions (experimental vs. control).

3.2 Results

In the discrimination task, a two-factor RM-ANOVA examining the effects of Talker Condition and Block found that there was no significant difference in overall preference rates across testing blocks, $F(1, 38) = 3.484$, $p = 0.070$, $\eta^2 = 0.084$, but there was a significant effect of Talker Condition, $F(1, 38) = 6.028$, $p = 0.019$, $\eta^2 = 0.137$, with listeners selecting baseline tokens as “easier to understand” at higher rates for experimental talkers (mean (M) = 55.45%) than controls ($M = 50.85\%$). When these data were analyzed for each Talker Condition, preference rates were significantly different from chance only in response to experimental talkers, $t(78) = 5.054$, $p < 0.0001$; control $t(78) = 0.572$, $p = 0.569$.

We also found a significant Talker Condition $\times$ Block interaction, $F(1, 38) = 6.013$, $p = 0.019$, $\eta^2 = 0.137$, with simple effects tests showing a significant effect of Talker Condition for trials in Block 3, $F(1, 38) = 12.27$, $p = 0.001$, $\eta^2 = 0.244$, but not in Block 1, $F(1, 38) = 0.416$, $p = 0.523$, $\eta^2 = 0.011$. Preference rates to experimental talkers in both blocks differed significantly from chance, however, Block 1: $M = 55.13\%$, $t(38) = 3.124$, $p = 0.003$; Block 2: $M = 55.77\%$, $t(38) = 4.068$, $p = 0.0002$, while neither rate for control talkers did so (p 's > 0.09). Therefore, listeners showed a significant preference for experimental talkers' baseline speech over feedback-perturbed speech in both blocks, but no reliable preference in control talkers in either stimulus set. Preference rates for each condition and each block are shown in Figure 6.

In the recognition task, a two-factor RM-ANOVA returned a significant difference in accuracy and reaction time (RT) between Talker Conditions, accuracy: $F(1, 38) = 13.504$, $p = 0.001$, $\eta^2 = 0.262$; RT: $F(1, 38) = 209.5$, $p < 0.0001$, $\eta^2 = 0.846$. Listeners were more accurate ($M = 97.7\%$) and faster ($M = 2.06$ s) for experimental talkers than to controls (accuracy: $M = 95.5\%$, RT: $M = 2.49$ s). There was no main effect of Speech Condition, however, accuracy: $F(1, 38) = 0.962$, $p = 0.333$, $\eta^2 = 0.025$; RT: $F(1, 38) = 0.584$, $p = 0.449$, $\eta^2 = 0.015$, and the Speech Condition $\times$ Talker Condition interaction was also non-significant, accuracy: $F(1, 38) = 0.585$, $p = 0.449$, $\eta^2 = 0.015$; RT: $F(1, 38) = 3.881$, $p = 0.056$, $\eta^2 = 0.093$.

Given that there are inherent differences in intelligibility across talkers and no significant differences were observed across the two feedback conditions, it is likely that our experimental talkers were simply more intelligible overall than those randomly assigned to the control group. Indeed, we examined each recording epoch separately and even at baseline, in which feedback conditions were identical between the experimental and control talkers, the Talker Condition effect was significant, accuracy: $F(1, 38) = 10.65$, $p = 0.002$, $\eta^2 = 0.219$; RT: $F(1, 38) = 146.4$, $p < 0.0001$, $\eta^2 = 0.793$. Based on this initial overall difference, we analyzed the data from experimental talkers separately to determine whether it showed any independent evidence of feedback affecting intelligibility. Paired samples t -tests on the accuracy and RT data did in fact show a small but significant difference in RT between trials with speech from Baseline versus Epoch 2, $t(38) = -2.676$, $p = 0.011$, with listeners responding more slowly to speech in Epoch 2 (2.09 s) than at Baseline ($M = 2.02$ s). Accuracy differences were not significant, $t(38) = -0.165$, $p = 0.870$.

3.3 Discussion

Our perceptual investigation of speech produced under normal conditions versus with initial CI simulation of feedback found that changes in intelligibility were subtle. The majority of effects were found in the more sensitive discrimination task, probing listeners' explicit judgments of intelligibility. Strikingly, listeners in this task favored speech produced at baseline, rather than speech produced with typical Lombard effects in rate, amplitude, and F0. The results from our recognition task provide converging evidence: Although there were no differences in recognition accuracy between speech from Baseline and Epoch 2, listeners responding to the experimental group were slower for CI-simulated speech recognition compared to baseline speech. It appears, therefore, that the speech produced with CI-simulated feedback is not more intelligible than normal speech; in fact, it is actually slightly more difficult to recognize.

However, the above study may not provide the most relevant test of intelligibility. Speakers have been shown to adjust their speech to specific listening challenges (Hazan & Baker, 2011; Hazan, Gryn timer, & Baker, 2016), and it is possible that the adjustments made under CI simulation result in an intelligibility benefit *for listeners in the same conditions*, rather than listeners in general. In other words, speakers may have adjusted in ways that resulted in increased intelligibility for themselves (through the CI simulation transformation), improving their ability to monitor their own production. This possibility is tested below.

4 Vcoded speech intelligibility

In order to determine whether the speech changes observed in response to CI-simulated feedback increased intelligibility in the speakers' specific acoustic conditions, we replicated the perceptual study above but applied the CI simulation to all of the stimuli. If speech changes improved the transformation-specific intelligibility, then listeners should prefer speech that was produced with those effects (i.e., in Epoch 2) and/or be able to recognize such speech more easily.

4.1 Method

4.1.1 Subjects.—Fifty-six native speakers of English (24 males, 32 females), none of whom participated in the previous study, were recruited from Trinity College. Participants' mean age was 20.1 years with a range of 18–25. Participants reported no history of hearing loss and passed an audiometric screening (calibrated Ambco 650A) with thresholds 30 dB HL between 500 and 8000 Hz on the day of the experiment.

4.1.2 Stimulus materials.—The same stimuli from Section 3.1.3 were used, with one difference: all tokens were passed through the same eight-channel CI simulation transformation that was used as a feedback perturbation for talkers in the experimental condition during Epoch 2 (see Section 2.1.2 above).

4.1.3 Design, procedure, and data analysis.—Other than the difference in stimuli and the new set of participants, the method was identical to that described in Section 3.1 above.

4.2 Results

In the discrimination task, a two-factor RM-ANOVA examining the effects of Talker Condition and Block found that there was once again no significant difference in overall preference rates across testing blocks, $F(1, 55) = 0.75$, $p = 0.389$, $\eta^2 = 0.014$, but there was a significant effect of Talker Condition, $F(1, 55) = 11.10$, $p = 0.0015$, $\eta^2 = 0.168$, with listeners selecting baseline tokens as “easier to understand” at higher rates for experimental talkers ($M = 54.12\%$) than controls ($M = 49.69\%$). Just as in Study 1, preference rates were significantly different from chance only in response to experimental talkers, $t(111) = 5.48$, $p < 0.0001$; control $t(111) = -0.38$, $p = 0.710$.

We also found a significant Talker Condition $\times$ Block interaction, $F(1, 55) = 12.35$, $p = 0.0009$, $\eta^2 = 0.183$, with simple effects tests showing a significant effect of Talker Condition for trials in Block 3, $F(1, 55) = 19.62$, $p < 0.0001$, $\eta^2 = 0.263$, but not in Block 1, $F(1, 55) = 0.13$, $p = 0.724$, $\eta^2 = 0.002$. Unlike Study 1, here only the rate in response to experimental talkers in Block 3 ($M = 56.47\%$) differed significantly from chance, $t(55) = 5.99$, $p < 0.0001$; responses to control talkers in both blocks and the first block of experimental talker trials did not (all other p 's > 0.07).

In the recognition task, a two-factor RM-ANOVA showed no significant differences in response accuracy across Talker or Speech Conditions (all p 's > 0.08). Differences were observed in RT, however, across Talker Conditions, with faster RTs in response to experimental talkers than to those in the control, $F(1, 52) = 57.81$, $p < 0.0001$, $\eta^2 = 0.526$, and Speech Conditions, with faster RTs in response to Baseline tokens than to those produced in Epoch 2, $F(1, 52) = 7.55$, $p = 0.008$, $\eta^2 = 0.127$. The Speech Condition $\times$ Talker Condition interaction in RT was non-significant, $F(1, 52) = 1.937$, $p = 0.170$, $\eta^2 = 0.036$. These results therefore replicate the general intelligibility advantage of experimental talkers over controls seen in Study 1 (for both accuracy and RT, opposed to the current findings' limitation to RT), but do not show evidence of an intelligibility benefit for tokens under conditions matching the listening context in this study (i.e., CI simulation).

4.3 Discussion

The effects of CI-simulated feedback on intelligibility were largely parallel with those observed in Study 1. It cannot be, therefore, that the changes observed under CI simulation resulted in condition-specific intelligibility benefits (e.g., Hazan & Baker, 2011; Hazan et al., 2016) that could not be detected in the prior clear-speech task. We must conclude, therefore, that while speech intelligibility did not change drastically as a result of CI-simulated feedback perturbation, it was nonetheless significantly reduced overall.

5 General discussion and conclusions

Our goal in this research was to evaluate the similarity of the Lombard effect and the effect of real-time CI simulation in terms of their broader, non-compensatory impact on speech production. While the similarity between the perturbation responses was unclear following supra-segmental acoustic analysis, the relative decrease in initial intelligibility with CI-

simulated feedback coupled with reduced F0 variation leads us to conclude that they bear little more in common than a superficial resemblance.

The motivation for speakers' responses to CI-simulated feedback therefore remains unclear. The constellation of changes to speech production do not compensate for errors introduced into acoustic feedback (Casserly, 2015), nor do they serve to increase intelligibility in other ways via relatively preserved supra-segmental mechanisms. One current model of real-time speech motor control includes a means by which speakers could disengage from their sensory feedback if it was particularly unreliable or error-ridden (Houde & Nagarajan, 2011). It is entirely possible that our speakers could be disengaging as Houde and Nagarajan predict, as well as perhaps dedicating cognitive resources elsewhere—to somatosensory-based control, for example, or to perceptual learning associated with processing CI-simulated signals.

Regardless of the explanation for this specific response to this specific global feedback perturbation, however, we believe the present set of results has important implications for the investigation and interpretation of feedback perturbation. Specifically, it appears that feedback responses are complex, and that complexity may be the rule rather than the exception when it comes to this kind of behavior.

The Lombard effect is complex; it involves direct error-reducing compensation in terms of voice amplitude, but as we have discussed, there are a host of other changes observed in this response as well. Those changes are not directly tied to compensation for feedback error, and they may not even all contribute to the intelligibility benefits that serve as a hallmark of Lombard speech (Cooke et al., 2014). The Lombard effect also depends to some extent on the communicative intentions of a speaker, rather than being an irresistible “reflex” as it has sometimes been characterized (Garnier et al., 2010; Lane & Tranel, 1971; Summers et al., 1988).

What we have seen so far is that the response to CI-simulated feedback is similarly complex. Although there remains much to be documented, speakers produce systematic, non-compensatory changes in vowel acoustic-phonetics (Casserly, 2015), coupled with supra-segmental changes that appear at first to be related to strategies that increase intelligibility, but which ultimately have no such effect. Instead, speakers' intelligibility decreases. There is also evidence that the response to CI-simulated feedback changes over time and with increasing experience; this interaction with perceptual learning constitutes a promising avenue for further exploration in the future.

Speakers' responses to vowel formant perturbations (Houde & Jordan, 1998; Purcell & Munhall, 2006), shifts in F0 cues (Burnett et al., 1998; Elman, 1981), voice-onset time shifts (Mitsuya, MacDonald, & Munhall, 2014), and alterations of fricative spectra (Casserly, 2011; Shiller, Sato, Gracco, & Baum, 2009) seem to be relatively simple in comparison. A feedback error is introduced and speakers generally act to compensate and reduce the error. This apparent simplicity may, however, be an artifact of the set of experimental methods adopted in this literature. Acoustic cues are manipulated in isolation, and responses are typically assessed in the same domains which were targeted in the manipulation (but see

Villacorta, Perkell, & Guenther, 2007). It is possible that changes are occurring in other domains which these analyses do not detect, or that responses depend on aspects of the task that have not been subject to manipulation (e.g., communicative intent). Some studies exploring these aspects have shown perturbation responses to be unaffected by changes in speaker intent, for example (Munhall, MacDonald, Byrne, & Johnsrude, 2009), while others find that methodological changes affecting predictability, for example, can have large qualitative effects (Behroozmand, Sangtian, Korzyukov, & Larson, 2016). In any case, we believe that more complex perturbations of speech feedback are called for, along with similarly complex assessments of behavioral change in response.

Acknowledgements

We would like to thank Dr Ken de Jong for guidance in acoustic analysis, Luis Hernandez for technical assistance with the portable vocoder equipment, David Ballenger for assistance in manuscript review, as well as Nijel Hill, Connor Jacobson, and Viridiana Medina for assisting in the collection and initial implementation of the intelligibility studies.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders [NIDCD R01-00111 and NIDCD T32-DC00012], and by the Trinity College Dean of the Faculty Office.

References

- Baker R, & Bradlow AR (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, 52(4), 391–413. [PubMed: 20121039]
- Beckford Wassink A, Wright RA, & Franklin AD (2007). Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics*, 35(3), 363–379.
- Behroozmand R, Sangtian S, Korzyukov O, & Larson CR (2016). A temporal predictive code for voice motor control: Evidence from ERP and behavioral responses to pitch-shifted auditory feedback. *Brain Research*, 1636(1), 1–12. [PubMed: 26835556]
- Boersma P, & Weenink D (2008). Praat: Doing phonetics by computer. Version 5.0.40.
- Burkholder RA, Pisoni DB, & Svirsky MA (2005). Effects of a cochlear implant simulation on immediate memory in normal-hearing adults. *International Journal of Audiology*, 44(10), 551–558. [PubMed: 16317807]
- Burnett TA, Freedland MB, Larson CR, & Hain TC (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, 103(6), 3153–3161. [PubMed: 9637026]
- Casserly ED (2011). Speaker compensation for local perturbation of fricative acoustic feedback. *Journal of the Acoustical Society of America*, 129(4), 2181–2190. [PubMed: 21476673]
- Casserly ED (2015). Effects of real-time cochlear implant simulation on speech production. *Journal of the Acoustical Society of America*, 137(5), 2791–2800. [PubMed: 25994707]
- Casserly ED, Pisoni DB, Smalt CJ, & Talavage TM (2011). A portable, real-time vocoder: Technology and preliminary perceptual learning findings. *Journal of the Acoustical Society of America*, 129(4), 2527–2527.
- Cooke M, Mayo C, & Villegas J (2014). The contribution of durational and spectral changes to the Lombard speech intelligibility benefit. *Journal of the Acoustical Society of America*, 135(2), 874–883. [PubMed: 25234895]
- Eisenberg LS, Shannon RV, Martinez AS, Wygonski J, & Boothroyd A (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*, 107(5), 2704–2710. [PubMed: 10830392]

- Elman JL (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America*, 70(1), 45–50. [PubMed: 7264071]
- Fu QJ, Nogaki G, & Galvin JJ (2005). Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology*, 6(2), 180–189. [PubMed: 15952053]
- Fu QJ, Shannon RV, & Wang X (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *Journal of the Acoustical Society of America*, 14(6), 3586–3596.
- Garnier M, Henrich N, & Dubois D (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language and Hearing Research*, 53(3), 588–608.
- Hazan V, & Baker R (2011). Acoustic–phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, 130(4), 2139–2152. [PubMed: 21973368]
- Hazan V, Gryn timer J, & Baker R (2016). Is clear speech tailored to counter the effect of specific adverse listening conditions? *Journal of the Acoustical Society of America*, 132(5), EL371–377.
- Houde JF, & Jordan MI (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. [PubMed: 9469813]
- Houde JF, & Jordan MI (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310.
- Houde JF, & Nagarajan SS (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5(82), 1–14. [PubMed: 21283556]
- Huggins a W., & Nickerson RS (1985). Speech quality evaluation using “phoneme-specific” sentences. *Journal of the Acoustical Society of America*, 77(5), 1896–1906. [PubMed: 3998299]
- Kaiser AR, & Svirsky MA (1999). A real time PC-based cochlear implant speech processor with an interface to the Nucleus 22 electrode cochlear implant and a filtered noiseband simulation In *Research on Spoken Language Processing: Progress Report* (Vol. 23, pp. 417–428). Bloomington, IN: Indiana University Retrieved from http://www.iu.edu/~srlweb/pr/PR23_p417_A_Real_Time_PC-Based_Cochlear_Implant_Speech_Processor_with_an_Interface_to_the_Nucleus_22_Electrode_Cochlear_Implant_and_a_Filtered_Noiseband_Simulation.pdf
- Lane H, & Tranel B (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 4(4), 677–709.
- Lindblom BEF (1990). Explaining phonetic variation: a sketch of H&H theory In Hardcastle HJ & Marchal A (Eds.), *Speech production and speech modeling*, NATO ASI Series D: Behavioural and social sciences (Vol. 55, pp. 403–439). Dordrecht, The Netherlands: Kluwer A.P.
- Loebach JL, & Pisoni DB (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *Journal of the Acoustical Society of America*, 123(2), 1126–1139. [PubMed: 18247913]
- Loebach JL, Pisoni DB, & Svirsky MA (2009). Transfer of auditory perceptual learning with spectrally reduced speech to speech and nonspeech tasks: Implications for cochlear implants. *Ear and Hearing*, 30(6), 662–674. [PubMed: 19773659]
- Mitsuya T, MacDonald EN, & Munhall KG (2014). Temporal control and compensation for perturbed voicing feedback. *Journal of the Acoustical Society of America*, 135(5), 2986–2994. [PubMed: 24815278]
- Munhall KG, MacDonald EN, Byrne SK, & Johnsrude IS (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *Journal of the Acoustical Society of America*, 125(1), 384–390. [PubMed: 19173425]
- Niparko JK (2009). *Cochlear implants: principles & practices*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Nusbaum HC, Pisoni DB, & Davis CK (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words In *Research on Speech Perception Progress Report No. 10* (pp. 357–376). Bloomington, IN: Indiana University Retrieved from http://www.academia.edu/403123/Sizing_up_the_Hoosier_Mental_Lexicon_Measuring_the_Familiarity_of_20_000_Words
- Patel R, & Schell KW (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, 51(1), 209–220.

- Pick HL, Siegel GM, Fox PW, Garber SR, & Kearney JK (1989). Inhibiting the Lombard effect. *Journal of the Acoustical Society of America*, 85(2), 894–900. [PubMed: 2926004]
- Purcell DW, & Munhall KG (2006). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America*, 119(4), 2288–2297. [PubMed: 16642842]
- Qin MK, & Oxenham AJ (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, 114(1), 446–454. [PubMed: 12880055]
- Reinfeldt S, Ostli P, Håkansson B, & Stenfelt S (2010). Hearing one's own voice during phoneme vocalization—transmission by air and bone conduction. *Journal of the Acoustical Society of America*, 128(2), 751–62. [PubMed: 20707445]
- Shafiro V (2008). Identification of environmental sounds with varying spectral resolution. *Ear & Hearing*, 29(3), 401–420. [PubMed: 18344871]
- Shafiro V, Sheft S, Gygi B, & Ho KTN (2012). The influence of environmental sound training on the perception of spectrally degraded speech and environmental sounds. *Trends in Amplification*, 16(2), 83–101. [PubMed: 22891070]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, & Ekelid M (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304. [PubMed: 7569981]
- Shiller DM, Sato M, Gracco VL, & Baum SR (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America*, 125(2), 1103–1113. [PubMed: 19206885]
- Stenfelt S, & Håkansson B (2002). Air versus bone conduction: An equal loudness investigation. *Hearing Research*, 167(1–2), 1–12. [PubMed: 12117525]
- Summers WV, Pisoni DB, Bernacki RH, Pedlow RI, & Stokes MA (1988). Effects of noise on speech production: acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, 84(3), 917–928. 10.1121/1.396660 [PubMed: 3183209]
- Villacorta VM, Perkell JS, & Guenther FH (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America*, 122(4), 2306–2319. [PubMed: 17902866]
- Vogel AP, Fletcher J, & Maruff P (2014). The impact of task automaticity on speech in noise. *Speech Communication*, 65(1), 1–8.

Figure 1.

Signal transformation used to simulate cochlear implant acoustic processing. Natural sound (upper left) is spectrally analyzed (upper right) to determine the signal strength within particular bandwidths of frequencies. Envelopes for each of these frequency “channels” are then applied to broadband noise (lower right), shaping it into a spectrally-degraded version of the original signal (lower right, spectrum; lower left, spectrogram).

Figure 2. Mean durations for sentences produced by speakers in the experimental ($n = 9$; solid line) and control ($n = 3$; dashed line) groups (color online).

Figure 3. Average F0 of sentences. F0 changed significantly across Epochs in the experimental group (solid line), but not in the control (dashed line) (color online).

Figure 4.

Average sentence F_0 variability (standard deviation) produced by female speakers in the experimental ($n = 7$; solid line) and control ($n = 3$; dashed line) groups. No change across epochs was observed in the control group, while experimental talkers' F_0 variability was significantly lower in Epoch 2 than in Epochs 1 or 3 (color online).

Figure 5. Mean voice amplitudes for sentences. Amplitudes changed significantly across Epochs in the experimental group (solid line), but not in the control (dashed line) (color online).

Figure 6.

Rates at which listeners chose speech tokens produced with normal acoustic feedback as “easier to understand” under acoustic transformation than speech from Epoch 2. Responses to experimental talkers in black (solid line) and to controls in gray (dash-dot line). Only judgments of experimental talkers’ speech differed significantly from chance (50%, dashed line).