Perceptual learning and sensorimotor learning with cochlear-implant simulated speech feedback

Abigail R. Bradshaw[1]*

Susan Black[1]

Clément Gaultier[1,2]

Matthew H. Davis[1]

[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, UK, CB2 7EF

[2]Hearing Institute, Institut Pasteur, 63 Rue de Charenton, Paris, France.

*Corresponding author information:
Dr Abigail Bradshaw, MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, UK, CB2 7EF (email: abbie.bradshaw@mrc-cbu.cam.ac.uk).

Funding

This work was funded by an Experimental Psychology Society Small Grant and a Newnham College Amy Whiteley Research Fellowship (both awarded to Abigail Bradshaw). Matthew Davis was supported by UK Medical Research Council funding of the Cognition and Brain Sciences Unit (MC_UU_00030/6).

Cochlear implants (CIs) provide deaf individuals with access to auditory feedback from their own voice during production. This experiment investigated whether typical hearing participants can use CI simulated speech feedback for perceptual learning and sensorimotor control of speech. CI simulation was achieved via noise vocoding, a technique that degrades the spectral detail in a speech signal in a manner similar to a CI. 32 participants took part in the experiment. First, participants were tested on their recognition of noise vocoded sentences before and after a training task; either perception training, where participants listened to noise vocoded sentences while reading matching text; or production training, where participants read aloud sentences whilst hearing their own voice noise-vocoded in real-time. Both groups of participants then underwent a speech motor adaptation paradigm in which formants were perturbed in real-time noise vocoded speech auditory feedback. Both perception and production training tasks resulted in significant improvements in recognition of noise vocoded sentences, with no effect of training type. Speech motor adaptation however was not significant at the group level in response to the formant perturbations. This suggests that successful perceptual learning for degraded speech is not sufficient for successful sensorimotor learning with degraded auditory feedback.

Successful communication with others requires both intelligible speech production and accurate speech perception. Both of these functions are affected by deafness, limiting the ability for deaf individuals to communicate with others through the oral modality. This lack of access to speech auditory feedback (the sound of one's own voice while speaking) can be associated with atypical development and maintenance of speech articulation. Congenitally deaf infants do not show the typical developmental stage of babbling, thought to facilitate learning of sensorimotor mappings (Oller & Eilers, 1988), and such individuals struggle to acquire typical speech articulation, making them less intelligible (Smith, 1975). Further, loss of hearing later in life (post-lingually) can result in subtle deterioration of speech (Cowie et al., 1982; Lane & Webster, 1991); although there is high variability in outcomes, related to factors such as age of onset of deafness.

Cochlear implants (CIs) are a sensory prosthesis that can restore hearing to deaf individuals, resulting in substantial improvements in their speech perception abilities (Boisvert et al., 2020). This is an impressive feat given the auditory distortions introduced by a CI, resulting in a loss of fine-grained spectral and temporal information which allows for only very coarse frequency analysis compared to that achieved by the biological cochlear. Achieving such improvements in perception however involves an initial period of rehabilitation over the course of weeks and months, as individuals adapt to the distorted auditory input provided by a CI (Svirsky et al., 2001; Tyler & Summerfield, 1996). The impact of implantation on speech production abilities has received comparatively less attention; however, studies suggest that speech becomes more intelligible, with changes in vocal pitch and loudness, and increased contrasts between consonants and separation between vowels (for a review, see Gautam et al., 2019). A study by (Menard et al., 2007) found that such improvements continued to increase over the first year following implantation, suggesting gradual learning processes; however, even at one year post-implantation, better performance was found with the implant turned on compared to off, suggesting the contribution of online auditory feedback to improved speech articulation.

Indeed, online processing of auditory feedback is proposed to play an important role in speech motor control (Guenther, 2016; Parrell et al., 2019; Parrell & Houde, 2019). This is demonstrated eloquently in the altered auditory feedback paradigm; here, real-time perturbations of acoustic features such as fundamental frequency (F0) or formants are found to result in implicit compensatory adjustments to speech productions that correct for the apparent sensory error (Burnett et al., 1998; Houde & Jordan, 1998). Such perturbations can be implemented randomly across a series of utterances, resulting in rapid within-utterance compensatory changes (Larson et al., 2007); or held constant across a period of speaking, resulting in a gradual learning response that builds up across time and persists once the perturbation is removed (i.e. shows after-effects), known as speech motor adaptation (Purcell & Munhall, 2006). This body of work demonstrates the central role of auditory feedback to allow for monitoring of the sensory consequences of speech movements online, and for sensorimotor mappings to be continually calibrated through offline updating according to prediction errors.

To date, there have been a small number of studies that have investigated responses to auditory feedback perturbations in individuals with CIs. Loucks et al., (2015) investigated compensation to random perturbations of F0 in 6 CI participants and 6 typical hearing participants. Responses across trials were first classified into compensatory or following, according to their direction of change (in the opposing or same direction as the perturbation, respectively). Interestingly, CI participants showed a larger magnitude of change for both compensatory and following responses compared to typical hearing participants, suggesting an atypical pattern of responses to perturbations. The separation of responses by direction however makes it difficult to conclude whether overall CI participants were successful in compensating for the perturbations. A more recent study by Gautam et al., (2020) reported that CI participants showed significant within-utterance compensatory responses to random F0 perturbations (without excluding any responses on the basis of their direction), but only when these perturbations were sufficiently large (more than 600 cents, as opposed to the more typical perturbation magnitude of 200 cents used in Loucks et al.). This study was further unusual in the high proportion of trials including a perturbation (70-80% as opposed to a typical proportion of 40-50%). Overall, inter-subject variability in responses was high, and weakly correlated with duration of implant use. Borjigin et al., (2024) investigated adaptation to sustained formant perturbations in CI participants, using a paradigm in which the magnitude of the perturbation was adaptively ramped up until a participant no longer showed further changes in formants; the perturbation was then held constant at that magnitude for a ‘hold’ phase. Adaptation in the CI group reached significance during the adaptive ramp phase of the experiment, but ceased to be significant during the hold phase. Compared to a group of typical hearing participants (speaking with perturbed clear speech feedback), CI participants showed reduced adaptation rate and magnitude, as well as smaller after-effects of learning.

Overall, the inconsistency in results across studies in this small body of work is likely a result of the fact that samples are often small in size (due to difficulties in recruiting from this population), and highly heterogeneous (e.g. in age of onset of deafness, time since CI implantation etc), making it difficult to draw conclusions about the effectiveness of CI speech auditory feedback in supporting speech motor control. In samples of CI users, it is also difficult to disentangle the effects of the degraded CI speech feedback from the effects of the experience of deafness itself; that is, weaker responses to auditory feedback perturbations in CI users could either reflect the insufficiency of degraded CI feedback for supporting compensation/adaptation, or the effects of a period of deprivation of auditory feedback caused by deafness prior to implantation (especially if this occurred early in life in a critical period for speech development).

An alternative approach is to use simulation of CI speech to test the effects of such an altered signal on speech in typical hearing participants. This allows for the recruitment of larger, more homogeneous samples, and for the isolation of the effects of degraded speech auditory feedback specifically whilst controlling for previous hearing experience. Simulation of the signal processing implemented by a CI can be achieved using a technique known as vocoding (Shannon et al., 1995). This implements a degradation of the spectral content of speech similar to that introduced by a CI. This has been used extensively to investigate the effects of such degradation on passive speech perception, by applying such a technique to offline recordings of speech (Davis et al., 2005; Hervais-Adelman et al., 2008; Sohoglu et al., 2014). By contrast, only a small number of studies have looked at the effects of a real-time simulation of hearing one’s own speech auditory feedback through a CI (Casserly, 2015; Casserly et al., 2018; Casserly & Marino, 2024), and none have combined this with a perturbation of formants or F0 to assess compensatory behaviour or speech motor learning. The present study aimed to use real-time vocoding of speech auditory feedback with typical hearing participants, to test whether they show speech motor adaptation to a formant perturbation under CI simulated speech feedback.

A second aim of the current study was to investigate potential interactions between experience of speaking with CI simulated speech feedback and subsequent perceptual performance when passively listening to recordings of CI simulated speech. There is a large literature documenting interactions between speech production and perception (e.g. Bradshaw et al., 2024; Murphy et al., 2023; Pardo et al., 2022; Skipper et al., 2017). Within CI populations, there is evidence that speech production abilities in childhood are significantly correlated with receptive speech and language skills both at the same time point (Blamey et al., 2001), and at future time points (Casserly & Pisoni, 2013). Gains in speech intelligibility with implantation may therefore result in gains in perceptual abilities. Experience of speaking with CI speech auditory feedback may provide input that can support the process of perceptual learning to facilitate better recognition of speech through the degraded input provided by a CI.

Previous work with typical hearing participants has shown that recognition accuracy for vocoded speech improves simply with repeated exposure to it, attributed to a process of perceptual learning (Davis et al., 2005; Rosen et al., 1999). Davis et al., (2005) further demonstrated that such an improvement can be accelerated when listeners have prior knowledge of the content of the distorted speech; for example through provision of the written form prior to hearing the distorted speech (Davis et al., 2005). Such written cues result in an immediate perceptual pop-out effect, in which previously unintelligible sentences become highly intelligible; additionally, they also lead to improved recognition for subsequent novel vocoded material when written cues are then removed, suggesting top-down facilitation of perceptual learning processes. Experience of speaking with vocoded speech feedback may provide a similar means of facilitating perceptual learning, via provision of prior knowledge and employment of forward modelling processes. Indeed, such experience may be expected to confer a larger facilitation of perceptual learning; for example, due to more optimal temporal integration of prior knowledge of the lexical content of speech with the incoming distorted signal. This experiment therefore additionally aimed to investigate whether experience of speaking with real-time vocoded speech feedback resulted in significant improvements in subsequent recognition of vocoded speech, and whether this was over and above improvements observed with written cues.

Methods

The design, hypotheses and analyses for this experiment were pre-registered prior to collection of data on the Open Science Framework (https://osf.io/9rf68). Pseudonymised data and analysis code are also available on this platform (https://osf.io/3xj5e/).

Participants

In total 35 typical-hearing participants (30 female, 4 male, 1 prefer not to say), aged between 18-40 years (mean age = 23.7, SD = 5.28) and with no reported history of speech, language, or reading difficulties, took part in this experiment. Data from two participants was excluded as on testing it was found that English was not their first language. A further participant was excluded due to issues with their audio recordings. The remaining 32 participants included in the final analysis all reported their first language as English, and all but one spoke British English (the remaining participant spoke American English).

Procedure and Apparatus

The experiment was divided into two phases (see Figure 1). Phase 1 was a vocoded speech recognition task, which tested participants' ability to report words in noise vocoded sentences before and after a training phase. Phase 2 was a speech sensorimotor learning task that tested participants' ability to adapt to a formant perturbation with real-time vocoded speech auditory feedback. Both phases were completed in a single experimental session lasting around one hour. Each participant was seated in front of a computer screen and keyboard in a sound proof booth. During both phases, participants wore a headset microphone (Shure WH20) at a distance of approximately 5cm from their mouth, and circumaural headphones (Beyerdynamic DT 770 PRO 80 Ohm).

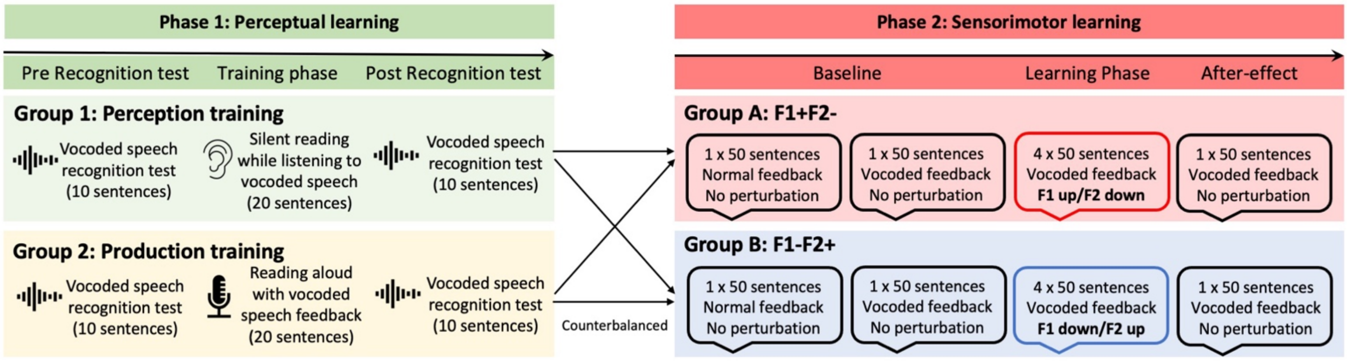
Phase 1: Perceptual learning
Pre Recognition test
Training phase
Post Recognition test
Group 1: Perception training
Vocoded speech recognition test (10 sentences)
Silent reading while listening to vocoded speech (20 sentences)
Vocoded speech recognition test (10 sentences)
Group 2: Production training
Vocoded speech recognition test (10 sentences)
Reading aloud with vocoded speech feedback (20 sentences)
Vocoded speech recognition test (10 sentences)
Counterbalanced
Phase 2: Sensorimotor learning
Baseline
Learning Phase
After-effect
Group A: F1+F2-
1 x 50 sentences Normal feedback No perturbation
1 x 50 sentences Vocoded feedback No perturbation
4 x 50 sentences Vocoded feedback F1 up/F2 down
1 x 50 sentences Vocoded feedback No perturbation
Group B: F1-F2+
1 x 50 sentences Normal feedback No perturbation
1 x 50 sentences Vocoded feedback No perturbation
4 x 50 sentences Vocoded feedback F1 down/F2 up
1 x 50 sentences Vocoded feedback No perturbation

Figure 1: Schematic illustration of the design of the experiment.

In phase 1, participants undertook a vocoded speech recognition test before and after a training task (pre- and post-tests). In each of these tests, participants listened to a series of 10 vocoded sentences through headphones, and after each were instructed to type out as much of the sentence as they could. Participants were randomly assigned to one of two training conditions; a perceptual training task and a production training task (16 participants per condition). For the perceptual training task, participants again listened to a series of vocoded sentences, but concurrent with the audio each sentence was also presented in written form on the screen. Participants were instructed to silently read the sentence while listening to it. For the production training task, participants instead saw each written sentence on the screen, and were instructed to read it aloud whilst hearing their voice vocoded in real-time through headphones. Each training task consisted of 20 trials. To assess participants' recognition of noise vocoded speech in the pre- and post-tests, a metric known as the token sort ratio was calculated using an online assessment tool (Bosker, 2021). This measure computes the orthographic similarity between two strings (here, the target sentence and the participant's response), assigning a score from 0 to 100. Briefly, words in the target sentence and the response are first sorted alphabetically, before a ratio is calculated to reflect the extent of shared substrings between the two. This method is more robust against the effects of misspellings than other automated methods, and has been shown to correlate highly with manual human scoring (Bosker, 2021).

In phase 2, both groups of participants from phase 1 underwent a speech motor adaptation paradigm. In this task, participants read sentences aloud whilst hearing their voice played back to them in real-time through headphones, with various manipulations at different stages of the task. Each trial began with visual presentation of the sentence to be read, which remained on screen for 4.5 seconds, with a 1 second inter-trial interval. Participants were instructed to read the sentence as soon as it appeared onscreen. Participants first underwent 10 practice trials with clear speech auditory feedback to familiarise themselves with the task, and practice speaking at a suitably loud level. This was aided by the use of visual feedback from an LED light display, calibrated such that a speaking level of 65-75dB (LAeq 1s) activated green lights, with speech louder or quieter activating red and yellow lights respectively. The level of speech playback through the headphones varied dynamically with changes in the amplitude of the participant's voice, but was calibrated such that the playback was amplified by about 5dB relative to their speaking level. In addition, during this task masking pink noise was played through headphones at 65dB; together with the amplification of speech feedback, this aims to mask perception of the participant's natural (unaltered) voice.

The main task was made up of 7 blocks of 50 trials, with the same set of 50 sentences presented in a random order in each block. For block 1, participants heard their voice played back through headphones without any manipulation (no vocoding or formant perturbation). From block 2 onwards, participants heard their voice vocoded (see stimuli for details). A perturbation of the first and second formants was gradually introduced in block 3 (ramp phase), ramping up over the first 25 trials of this block before being held constant for a further 3 blocks (blocks 4-6, hold phase). The perturbation was then removed for the final block (block 7, after-effect phase). The direction of this formant perturbation was manipulated between-subjects, to be either an upward shift of F1 and downward shift of F2, or a downward shift of F1 and upward shift of F2. For both groups, F1 and F2 were perturbed by 49.5 mels each (in their respective directions), resulting in a combined perturbation of 70 mels in F1/F2 space. The assignment of participants to each perturbation condition was counterbalanced with respect to their assignment to training groups in Phase 1 of the experiment.

The formant perturbation was implemented by an openly available MATLAB-based software application, Audapter (Cai, 2015; Cai et al., 2008). Recording, real-time processing and playback of speech was achieved using an audio interface (RME Fireface) and a mixer (Behringer). Speech was recorded at a sampling rate of 48 kHz (down-sampled to 16 kHz) with a buffer size of 96 samples. The total feedback loop latency of the set-up was measured at 20ms, according to the methods outlined in (Kim et al., 2020). This latency is well below the delay levels previously reported to disrupt speech adaptation (Max & Maffett, 2015; Shiller et al., 2020).

Stimuli

All sentence stimuli for both phases of the experiment were taken from the Harvard IEEE corpus of sentences (IEEE Subcommittee on Subjective Measurements, 1969). The vocoded stimuli for phase 1 were created using recordings of a female speaker of standard Southern British English reading a set of 40 of these sentences. From these 40, two sets of 10 sentences were used for the pre- and post-test tasks, with the order of these sets being counterbalanced across participants. The remaining 20 sentences were used for both the production and perception training tasks (with only the written form used for production training). A further set of 50 sentences were selected from this corpus for use as written stimuli for prompting speech productions in Phase 2. The same 50 sentences were repeated across blocks in a random order.

Vocoding of speech recordings for use in Phase 1 and of real-time speech in Phase 2 was achieved using a modified version of the software application Audapter (Cai, 2015; Cai et al., 2008). This software is designed to perform online modifications to speech in real-time (as in Phase 2 of this experiment), but can also be applied to pre-collected speech recordings ‘offline’, by feeding in the pre-recorded speech signal frame by frame. Our modified version incorporates custom code to implement spectral degradation of speech via noise vocoding as follows. First, a fast Fourier transform of the frame of speech is taken, and a noise signal of equal frame length to the speech is generated. A moving average is then computed over frequency on the phase and magnitude of the speech signal separately, using a user specified window size. Larger window sizes will result in averaging over wider frequency ranges (akin to averaging values across a larger number of pixels in blurring an image), resulting in a more degraded signal. Varying this parameter can be considered conceptually similar to varying the number of channels in a cochlear implant. The resulting subsampled speech Fourier transform is then multiplied by the noise Fourier transform (by multiplying magnitudes and adding phases). The real and imaginary parts are then recombined, before an inverse Fourier transform is performed to generate a degraded signal for speech feedback.

The value chosen for the window size parameter for both phases of the experiment (i.e. for vocoding of both pre-recorded stimuli and real-time speech) was roughly equivalent to the spectral degradation introduced by an 8 channel vocoder. This level was chosen based on an online pilot study (*n* = 10) which ran the perceptual training condition of phase 1; this found an average score (token sort ratio) of 72% for reporting words in vocoded sentences in the pre-test phase, comparable to findings on word report for sentences in quiet with CI participants (74%) as reviewed in (Boisvert et al., 2020).

Acoustic analysis

A custom Praat (Boersma & Weenink, 2021) script was used to track formants in speech recordings collected from phase 2 of the experiment. This first isolates the vocalised portions of speech using Praat’s autocorrelation method (Boersma, 1993), before extracting F1 and F2 values in Hertz using a Linear Predictive Coding approach. These were averaged across each sentence by taking the mean, and then converted into mels.

Quantification of adaptation

Adaptation to the formant perturbation was measured for each participant via two outcome variables: a production change measure and a vector of adaptation measure. Both rely on comparisons of formant frequencies between block 2 (the baseline block with vocoded speech feedback immediately prior to the introduction of the formant perturbation) and the subsequent blocks, with a particular focus on block 6 (the final block with altered feedback).

Firstly, a production change measure was calculated for F1 and F2 separately in which each of the 50 produced formant frequencies in block 6 (the final adaptation block) was normed to the F1/F2 frequencies for block 1 (baseline) on a sentence-by-sentence basis; the average of these values was then taken to give an average production change value for each participant (for each formant).

Secondly, a vector of adaptation measure was calculated, which quantifies the extent to which these changes in produced F1 and F2 values directly counter the direction of the feedback perturbation in F1-F2 space. To do this, firstly the inverse of the vector representing the feedback shift in F1-F2 space experienced by that participant was found; this vector represents perfect compensation to the feedback perturbation. The angular difference between this inverse shift vector and a vector representing the participant’s production change (relative to block 2) was then calculated; the cosine of this difference was then multiplied by the magnitude of production change. This vector of adaptation thus quantifies the degree to which the observed change in produced formants (i.e. the production change measure above) precisely opposed the feedback perturbation. This measure was calculated for each individual trial and then averaged within each block after the introduction of the feedback perturbation (blocks 3-7). Note that since this measure takes into account the direction of the formant perturbation experienced by the participant, positive values for both perturbation conditions represent opposing responses, while negative values indicate a following response (moving in the same direction as the perturbation).

Our pre-registered exclusion criteria were that whole datasets from a participant would be excluded if they made significant speech errors on more than 20% of trials in either block 2 or block 6 (the main blocks of interest for measuring adaptation). No participants exceeded this criterion, and so all were kept in our analyses. Two participants did not contribute data for block 7 of the adaptation task due to a technical issue with the experiment.

Hypotheses

For phase 1, we predicted that improvement in recognition of noise vocoded speech from pre- to post-training would be significantly greater in the production training group compared to the perception training group. For phase 2, we predicted that both perturbation groups would show significant adaptation, by moving their produced formant frequencies in an opposite direction to the direction of the perturbation they experienced.

All statistical analyses reported here were pre-registered, unless labelled as exploratory.

Results

Phase 1: Perceptual learning

Accuracy (token sort ratio scores) for reporting words in noise vocoded sentences before (pre-test) and after (post-test) training are plotted in Figure 2 for each training group. To test whether improvement in recognition accuracy was greater in the production training group than the perception training group, a linear mixed effects model was run on token sort ratio scores using the lmerTest package in R (Kuznetsova et al., 2017). Two random intercept models were run with fixed effects of phase (pre-test and post-test) and training group (perception and production); one in which these had additive effects, and one in which they had interactive effects. Both models also included random intercepts of participant and target sentence. Random slopes were not included, since this resulted in singular fit. A likelihood ratio test found that the interactive model did not provide a better fit to the data than the additive model ($\chi^2(1) = 0.049$, $p = .826$). The additive model found a significant effect of phase, with greater accuracy at post-test compared to pre-test (β = -9.65, $t(588) = -9.06$, $p < .001$), but no significant effect of group. Follow-up contrasts using estimated marginal means found that both groups showed a significant increase in accuracy from pre-test to post-test ($p < .001$ in both cases).

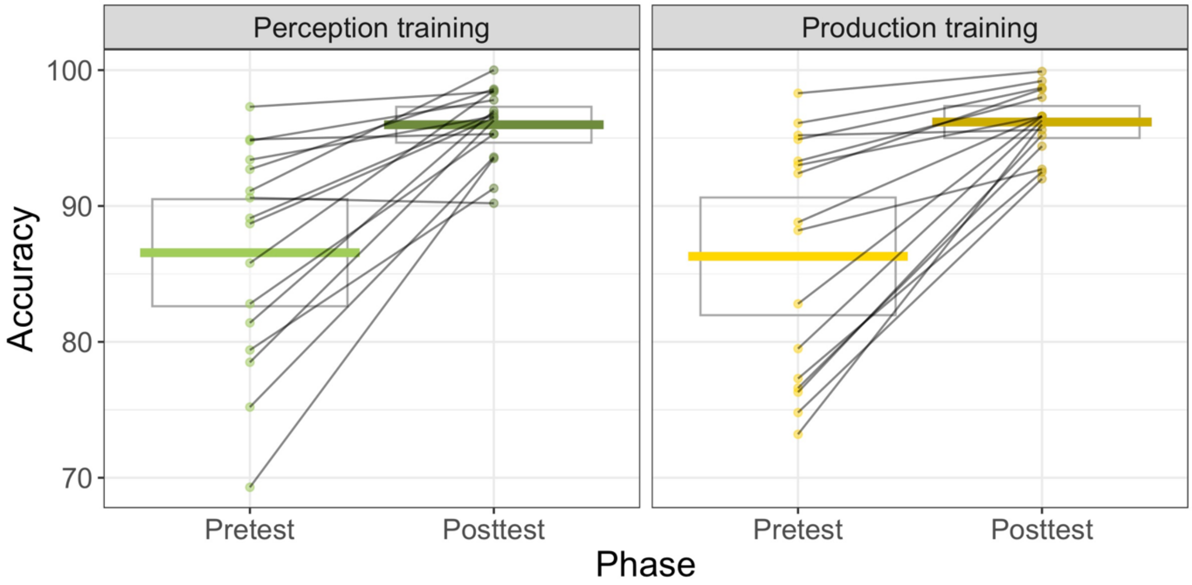

Figure 2: Accuracy (token sort ratio scores) for reporting words in noise vocoded sentences at pre-test and post-test in the two training groups. Dots indicate individual participants, thick lines indicate means, boxes indicate 95% confidence intervals.

Phase 2: Sensorimotor learning

Baseline changes: Effect of noise vocoding on speech

To test if the experience of speaking with real-time noise vocoded feedback (prior to introduction of the formant perturbation) had an effect on participants' produced formants, participants' average F1 and F2 values were compared across block 1 and block 2 by means of paired-samples t-tests. These changes are illustrated in Figure 3. Across all participants, there was a significant increase in both F1 ($t(1599) = 13.59$, $p < .001$) and F2 ($t(1599) = 1.9982$, $p = .046$) from block 1 to block 2.

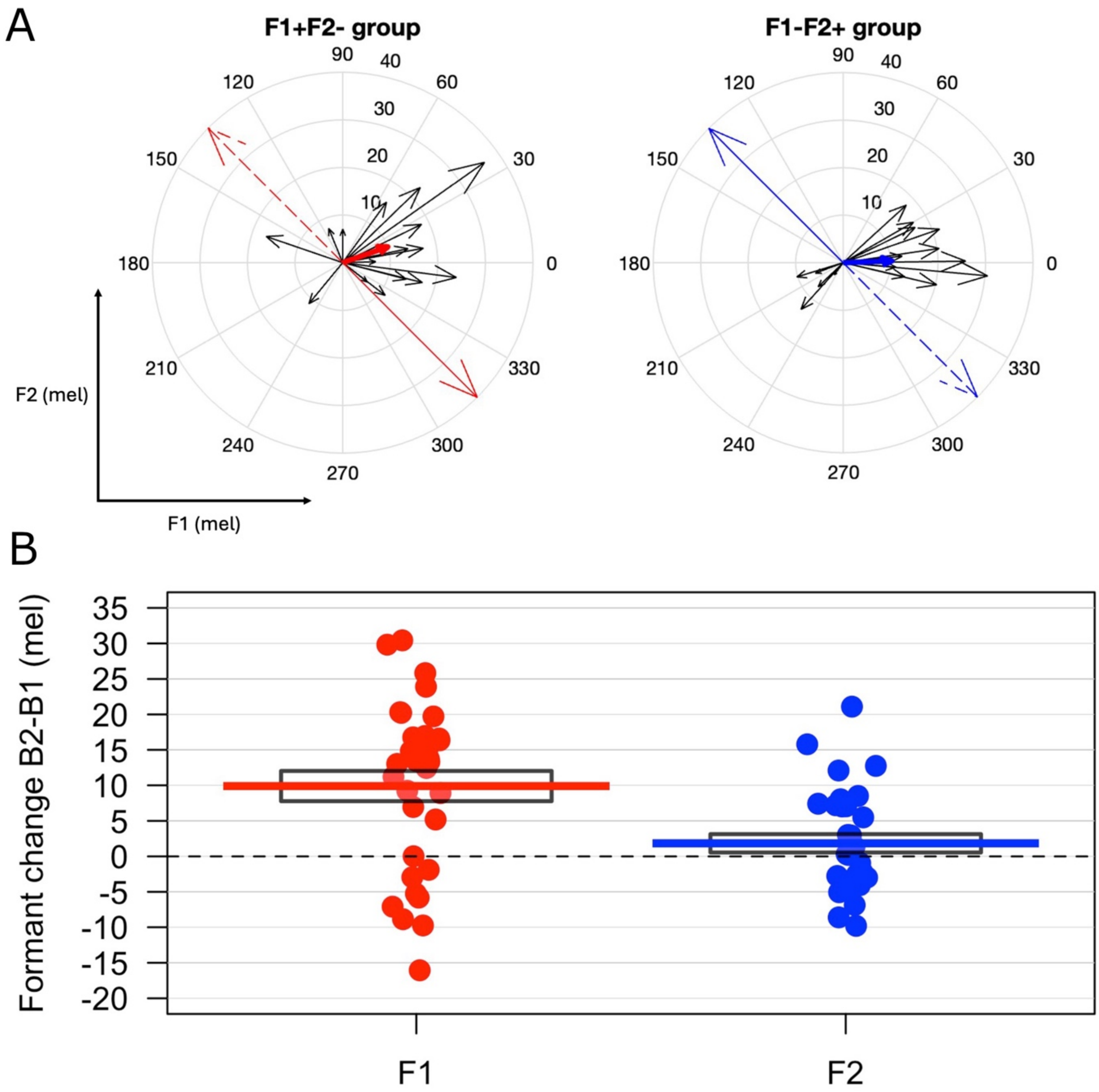

Figure 3: Changes in produced formants (in mels) from block 1 to block 2, reflecting changes associated with noise vocoded speech feedback (prior to introduction of the perturbation). (A) Changes plotted as vectors in formant space (for each perturbation group separately). Black arrows show individual participant data, thick red and blue arrows show group means. Thin red and blue arrows show direction of subsequent formant perturbation (solid arrow) and direction of hypothetical adaptation (dotted arrow) for reference purposes, although note that no formant perturbation was present at this stage of the experiment. (B) Changes in formants from block 1 to

block 2 plotted across all participants. Dots show individual participant averages, thick lines show group means, boxes show standard errors.

Adaptation changes: Effect of formant perturbation on speech

Changes in produced formants from noise-vocoded baseline (block 2) to the final block with formant perturbed noise vocoded feedback (block 6) are illustrated in Figure 4. To test the significance of adaptation at the group level, two-sided one-sample t tests were used on F1 and F2 production changes from Block 2 to Block 6 in each perturbation group. These found no significant changes in either F1 or F2 in either group. In an exploratory analysis, we further tested whether formant changes in block 6 were significantly different between the perturbation direction groups. We ran an LMM analysis on F1 and F2 production changes from block 2 to block 6, with fixed effects of perturbation direction group and formant, and random effects of sentence and participant. Random slopes were not included due to singular fit. A likelihood ratio test found a model with an interaction between group and formant provided a better fit to the data than a model in which these had additive effects ($\chi^2(1) = 8.36$, $p = .004$). Follow-up contrasts with this interactive model found a significant group difference for F1, in which F1 changes were significantly lower for the F1+F2- group than the F1-F2+ group ($\beta = -5.95$, $t(47.5) = -2.19$, $p = .034$). No significant group difference was found for F2 changes.

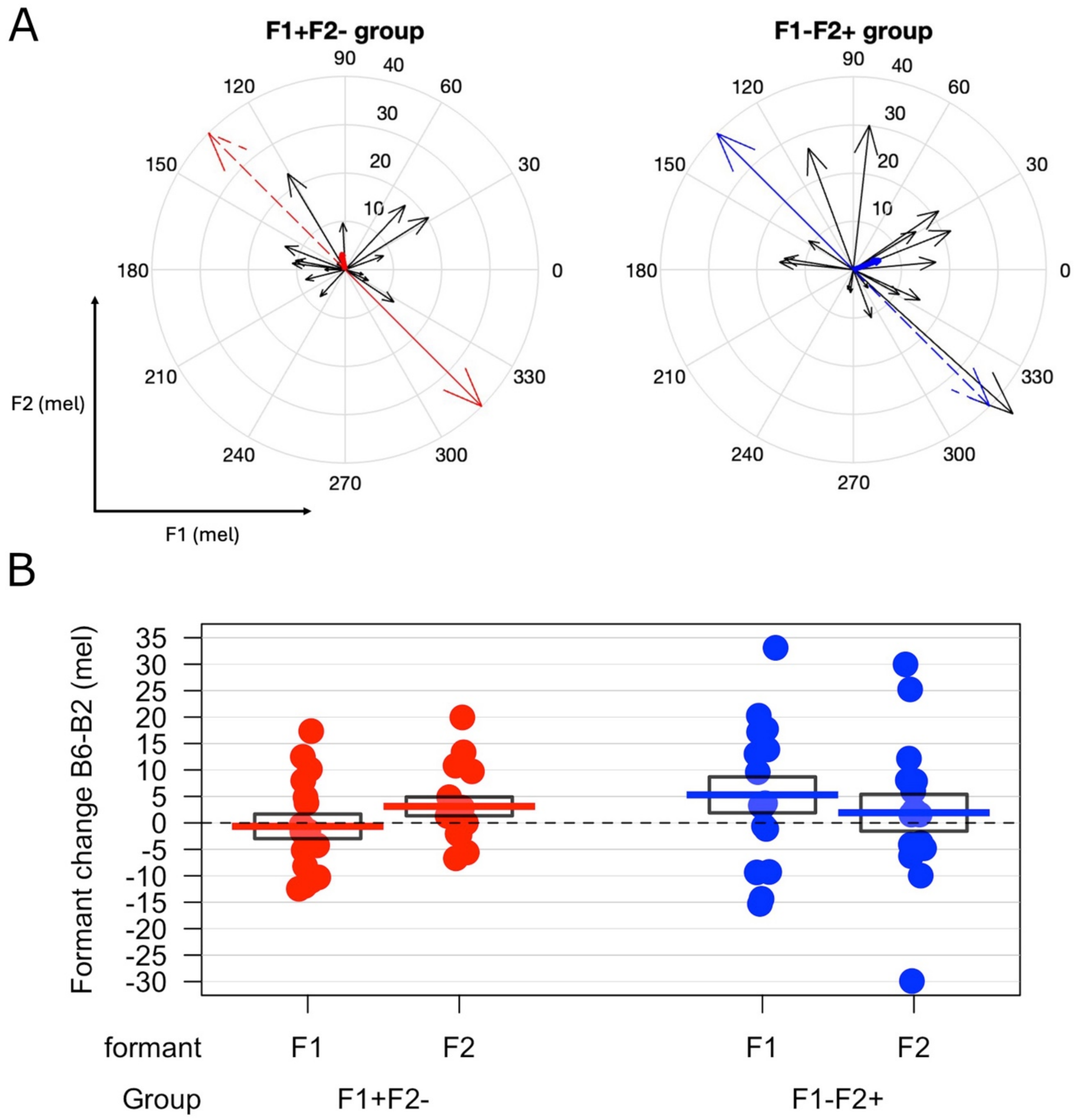

Figure 4: Changes in produced formants from block 2 (baseline) to block 6 (final block of formant perturbed feedback) in each perturbation group. (A) Changes plotted as vectors in formant space. Black arrows show individual participant data, thick red and blue arrows show group means. Thin red and blue arrows show direction of formant perturbation (solid arrow) and direction of hypothetical adaptation (dotted arrow), for reference purposes. (B) Production change in F1 and F2 in block 6 normalised to baseline block 2, for each perturbation direction group.

Dots show individual participant averages, thick lines show group means, boxes show standard errors.

To quantify to what extent changes in produced formants directly opposed the direction of the perturbation, a vector of adaptation measure was calculated. This is plotted for blocks 3 to 7 in Figure 5. To test whether each participant showed significant adaptation, we ran two-sided one-sample t-tests on each participant's vector of adaptation values from block 6, to classify participants as showing either a significant adaptation response (adaptation significantly greater than zero), a significant following response (adaptation significantly lower than zero), or no significant change. The number of participants in each of these categories is shown for the two groups in Table 1.

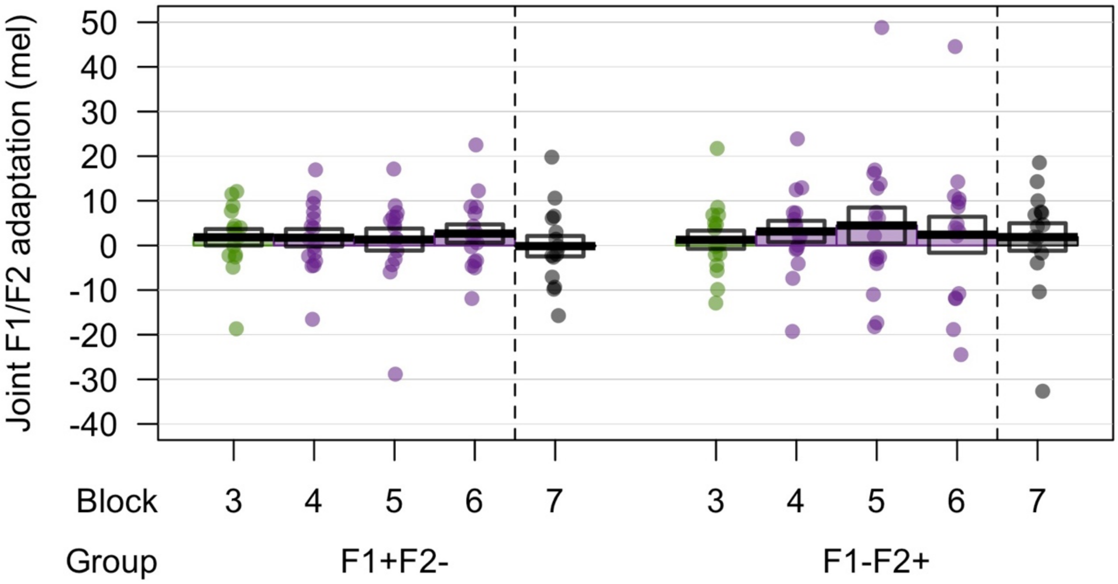

Figure 5: Vector of adaptation measure plotted across blocks 3-7 of the experiment, for each perturbation group. Dots show individual participant averages, thick lines show group means, boxes show standard errors.

Table 1: Frequency of participants showing adaptation, following or no change in formants in the two perturbation groups.

	F1+F2- group	F1-F2+ group
Significant adaptation response	2	4
Significant following response	1	4
No significant change	13	8

To investigate how adaptation changes across the blocks of altered feedback, a linear mixed effects model analysis was run on the vector of adaptation measure from blocks 3 to 7. We compared two models; one with a fixed effect of block (3-7) and random intercepts of sentence and participant; and an identical one with the addition of a fixed effect of perturbation direction group (F1+F2- or F1-F2+). Random slopes were not included due to failures of model convergence. A likelihood ratio test found the model containing fixed effects of block and group did not provide a better fit to the data than the model containing a fixed effect of block only ($\chi^2(1) = 0.058$, $p = .809$), supporting our prediction that this measure of adaptation would not be significantly different between the two groups (since this measure takes into account the direction of the perturbation experienced). Follow-up contrasts with the model containing a fixed effect of block found no significant differences between any of the blocks ($p > .2$ in all cases, using the Tukey method for adjusting for multiple comparisons). This suggests that there was no build up of adaptation across the blocks of formant perturbed feedback.

To explore whether training group in Phase 1 affected the magnitude of adaptation observed in Phase 2, an exploratory analysis was run to compare adaptation between production and perception training groups (collapsing across perturbation direction groups, see Figure 6). Those who experienced the production training task in Phase 1 had more exposure to speaking with noise vocoded feedback (an extra 20 sentences), potentially facilitating sensorimotor learning with this novel feedback. The adaptation vector measure was compared between groups for block 6 (the final block of altered feedback) by means of an independent-samples t-test. Despite a trend suggesting greater adaptation in the production training group, this group difference was not significant (*t*(26.71) = -1.74, *p* = .093). Changes in F1 and F2 at baseline from block 1 to block 2 further did not significantly differ according to which training condition participants completed in Phase 1 (F1: *t*(27.38) = -1.46, *p* = .157, F2: *t*(29.95) = -0.69, *p* = .49).

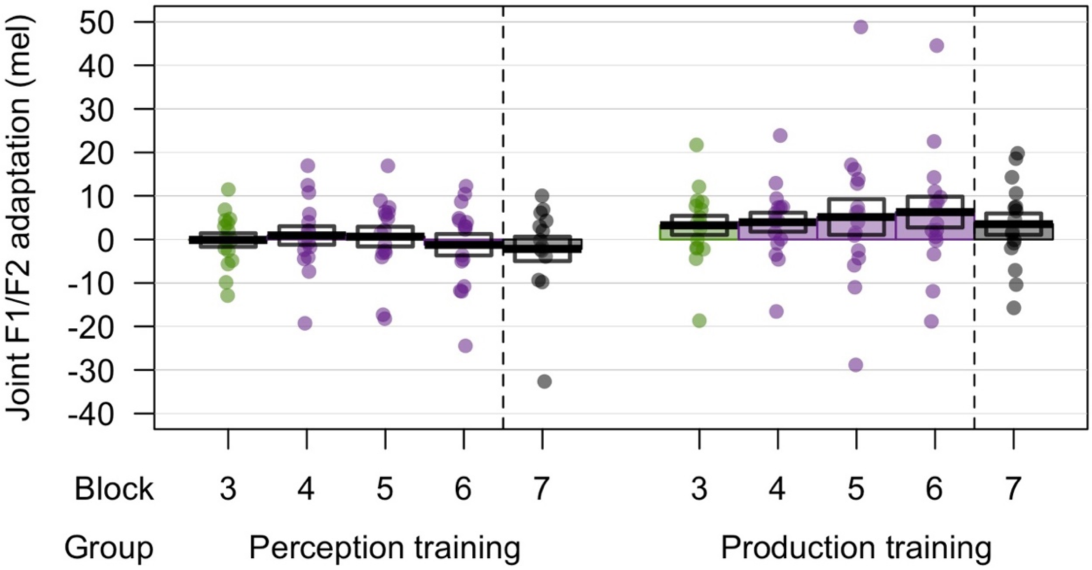

Figure 6: Vector of adaptation measure plotted across blocks 3-7 of the experiment, split by training group from phase 1 (collapsing across perturbation direction

conditions). Dots show individual participant averages, thick lines show group means, boxes show standard errors.

Discussion

This study investigated perceptual learning and sensorimotor learning with CI simulated (noise-vocoded) speech in typical hearing participants. In Phase 1, we found a significant improvement in recognition accuracy for noise vocoded sentences after both a perceptual training task (where listening was accompanied by written cues) and a production training task (where participants read aloud sentences whilst hearing their voice noise vocoded in real-time). Training task had no effect on the magnitude of this improvement, suggesting both tasks are equally effective at enhancing perceptual learning. In Phase 2, speaking with noise vocoded speech feedback resulted in significant increases in produced F1 and F2 frequencies, compared to speaking with clear speech feedback. However, when a formant perturbation was subsequently introduced to this noise vocoded speech feedback, we found no evidence of changes to formants indicative of speech motor adaptation. Overall therefore, the results suggest that successful perceptual learning for degraded speech is not sufficient for successful sensorimotor learning when that same degradation is applied to real-time speech auditory feedback during production.

In Phase 1, contrary to our predictions, the production training task was not associated with a significantly greater improvement in recognition of noise vocoded sentences compared to the perception training task. One concern could be the potential for ceiling effects at post-test, which may have limited the opportunity for further improvements in recognition. The level of noise vocoding used was chosen based on an online pilot experiment, in which average accuracy in the pre-test was found to be 72%; this is comparable to accuracy for sentence report in quiet in CI participants (Boisvert et al., 2020). Average accuracy at pre-test in our in-person study was found to be higher than this (around 86% for both groups), meaning that average accuracy at post-test approached (but did not reach) 100% (around 96% in both groups). Even so, the range of scores in our sample does not suggest that all participants were at ceiling at the post-test (ranging from 90.2% to 100%).

Nevertheless, this study demonstrates that experience of speaking with noise vocoded feedback is at least as effective at enhancing perceptual learning as the established effect of prior lexical knowledge during passive listening (Davis et al., 2005). Experience of speaking with CI simulated speech auditory feedback can therefore lead to gains in passive perception of such degraded input. This suggests that, for those CI individuals who wish to communicate in the oral modality, increased experience of speaking and hearing one's voice through a CI is likely to lead to improvements in passive perception of speech produced by others.

It is further interesting to consider a potential counter-argument to our original prediction, that instead we might have expected to see reduced perceptual learning during speech production due to speech-induced suppression; the phenomenon in which auditory input that is self-produced is perceptually and neurally suppressed relative to when that same auditory input is passively listened to (Heinks-Maldonado et al., 2005; Merrikhi et al., 2018; Ozker et al., 2024). Such suppression could in theory have hindered perceptual learning; instead, the robustness of the effect we observed with production training suggests that speech auditory feedback during production is sufficiently strongly represented and processed to support enhanced perceptual learning. It should be noted however that while auditory input doesn’t have to necessarily match the expected sensory consequences of a speech movement to be suppressed (with neural responses to tones presented during a period of speech motor planning showing perceptual and neural suppression (Max & Daliri, 2019; Merrikhi et al., 2018)), suppression is attenuated when speaking with altered auditory feedback (Behroozmand et al., 2015; Chang et al., 2013; Tourville et al., 2008). The unexpected noise vocoded feedback used here is thus likely to be associated with less suppression.

The observation of increases in produced F1 and F2 with noise vocoded feedback in the baseline blocks of Phase 2 is consistent with previously reported acoustic changes associated with Lombard effects; the tendency for speakers to involuntarily increase their vocal effort when speaking in noise (Lane & Tranel, 1971). Masking noise was present throughout both clear and noise vocoded baseline blocks (and for the rest of the adaptation task); however, the switch to noise vocoded speech feedback at baseline block 2 introduces an additional form of ‘noise masking’, which may have further exacerbated Lombard effects. Previous work has reported increases in F1 and F2 when speaking with noise masking (Castellanos et al., 1996; Van Summers et al., 1988). A previous study by Casserly (2015) looking at the effect of noise vocoded speech auditory feedback on word production in typical hearing participants similarly reported changes in produced F1; however, effects were vowel specific and consistent with a collapsing of vowel height contrast, rather than a global increase. This discrepancy may be due to the use of connected sentence stimuli in the current study, as opposed to isolated words.

Interestingly, formants did not appear to continue to increase across the remainder of the experiment after introduction of the formant perturbation. However, neither did they show the expected changes in opposition to the direction of the formant perturbations that would be indicative of adaptation. An exploratory analysis did find a significant difference between the perturbation direction groups in F1 changes, in which the group who experienced a downward shift of F1 showed greater increases in F1 than the group who experienced an upward shift of F1; a pattern consistent with adaptation. If the formant perturbation had no effect on production, we would expect no group differences according to the direction of the perturbation. This observed group difference is therefore interesting; however, it is difficult to interpret in the context of no significant changes in formants from baseline in either group. Across both groups, only 6 out of 32 participants showed evidence of a significant adaptation response, with nearly as many (5) showing a following response; that is, moving their formants in the same direction as the perturbation.

Following responses have generated much discussion in the speech sensory perturbation literature, and their underlying mechanism, as well as whether to exclude them from group-level analyses, remains debated. They have been attributed to an externalisation of the altered auditory feedback as not being self-generated (thus acting as an external referent to be matched rather than an error to be corrected) (Franken et al., 2023; Hain et al., 2000; Patel et al., 2014); a result of the state of the speech production system at perturbation onset (Franken, Acheson, et al., 2018); or simply the tail end of a unimodal distribution (Miller et al., 2023). In Miller et al (2023), data from across 22 studies of altered auditory feedback during single word production was pooled, and statistical tests performed to establish whether the distribution of compensation responses was unimodal or bimodal. This found clear evidence of a unimodal distribution, suggesting that such following responses do not represent a qualitatively different response type, but simply the tail end of a unimodal distribution. Sentence level adaptation tasks appear to be associated with both fewer following responses, and adaptation responses of greater magnitude than those observed during word production (Bradshaw et al., 2023; Lametti et al., 2018; Shiller et al., 2023). This would thus be consistent with the idea of a unimodal distribution shifted in the positive direction (meaning that fewer observations end up on the wrong side of zero). While it is therefore tempting to interpret the increased incidence of following responses in the present study as indicative of engagement in a qualitatively different sensorimotor process during speaking with noise vocoded feedback, it is perhaps more likely that these simply reflect the left-hand tail of a unimodal distribution centred on zero.

There are multiple possible explanations for why we didn’t observe significant adaptation at the group level with noise vocoded feedback in this study. One possibility is that participants did not perceive the noise vocoded speech feedback as self-generated, and so did not engage self-monitoring processes that support the detection and correction of sensory errors (Franken et al., 2023). Noise vocoded speech is a highly alien signal, that does not plausibly sound as if it can be generated by a human vocal tract. Sense of agency over a voice does appear to be surprisingly flexible, being maintained even in the face of dramatic alterations of speech auditory feedback, such as shifting its pitch by an octave (Franken et al., 2021), replacing it with another voice (Zheng et al., 2011), or even replacing it with one’s own voice speaking a different word (Lind et al., 2014). While the noise vocoded feedback was congruent in its timing and content with the participant’s natural voice, the degradation strips the acoustic signal of many cues to vocal identity, making it near impossible to determine speaker identity or even gender. It is possible therefore that this caused source-monitoring processes to reject the speech feedback as self-generated, resulting in reduced compensation to formant perturbations. If this was the case however, we might have expected to see a significant following response at the group level; perception of other voices typically engages phonetic convergence processes, in which we (unconsciously) adjust our voice to become more similar to that of the other speaker (Aubanel & Nguyen, 2020; Bradshaw & McGettigan, 2021; Pardo et al., 2017). Such convergence has been observed even when interacting with an artificial agent with a synthetic voice in human-computer interactions (Gessinger et al., 2021); however, it is possible that the noise vocoded speech used in the present study was too artificial to be recognised as another speaker.

An alternative possibility is that the degradation achieved by the noise vocoding process meant that the brain could not perceive or detect the formant perturbation. This should not be confused with explicit detection; the magnitude of formant perturbations typically used in altered auditory feedback experiments with clear speech are usually not detected by participants, and studies with pitch perturbations suggest that participants compensate regardless of whether they are explicitly aware of the perturbation or not (Franken, Eisner, et al., 2018; Hafke, 2008). There is evidence to suggest however that better auditory discrimination of formant shifts during passive perception is associated with greater adaptation to formant perturbations during production (Villacorta et al., 2007). The magnitude of the perturbation of formants used in the present study was relatively small (70 mel change across both F1 and F2); the coarse frequency information available in the degraded auditory feedback may therefore have simply made it too difficult for the brain (implicitly or otherwise) to detect this change. The results from Phase 1 suggests that these stimuli were highly intelligible to participants, with vowels being clearly perceived; however, it is possible that a larger formant perturbation may have been required to elicit adaptation with this level of degradation.

These results with CI simulated speech in typical hearing participants are somewhat discrepant with previous findings on speech motor adaptation in CI users. Borjigin et al., (2024) measured CI participant's adaptation to a perturbation of F1, and found evidence of significant adaptation at the group level. However, this adaptation was smaller in magnitude compared to that shown by typical hearing participants (with clear speech feedback), and was only significant during an adaptive ramp phase in which the magnitude of the perturbation was gradually increased in a manner sensitive to participant's adaptation behaviour. This adaptive ramp procedure resulted in a perturbation magnitude of more than 171Hz (218.82 mels), significantly greater than the perturbation magnitude used in the current study (joint 70 mel shift across F1 and F2). Unlike the typical hearing participants however, adaptation then ceased to be significant in the CI group during a hold phase, in which the magnitude of the perturbation was held constant at the maximum level at which participants stopped showing further changes. This finding appears more in line with the results found here, of no significant adaptation during a hold phase in typical hearing participants with CI simulated auditory feedback.

Taken together, the findings of ours and Borjigin et al's studies suggest that adaptation with CI feedback may only be possible when using an adaptive ramp procedure, and/or when employing a relatively large perturbation. It is also worth highlighting that while Borjigin et al., used single word production (with only one word stimulus of “Ed”), this study employed perturbations during variable sentence-level speech. While the existing literature with sentence-level adaptation tasks suggests that they may be associated with greater adaptation than word production tasks (Bradshaw et al., 2023; Lametti et al., 2018; Shiller et al., 2023), it is possible that for degraded speech, repeated production and experience of perturbed feedback for a single vowel sound may have been more effective at inducing adaptation, perhaps by allowing the perturbation to be more easily perceived.

It is also possible that longer exposure to the degraded speech feedback may be required for successful speech motor adaptation. However, the CI users in Borjigin et al. had between 5 and 55 years of experience of CI use, in sharp contrast to the minutes of exposure to CI simulated speech experienced by our typical hearing participants. This suggests that, even with prolonged experience of hearing one’s voice through a CI, sensorimotor learning with this degraded speech auditory feedback remains atypical in deaf individuals. It would be of interest to study changes across time following implantation on measures of sense of agency over speech auditory feedback, speech intelligibility and sensorimotor learning, to see how these may change and interact with one other with increasing experience of this new degraded self-voice. This could be complemented by studies employing CI simulation with typical hearing participants over longer timescales than those employed in the present study (e.g. exposure across days or weeks), to try to tease apart the effects of degraded speech feedback from experience of deafness.

One further point worth highlighting in the present study is the use of background masking noise during speech production blocks in Phase 2 of the experiment. The use of masking noise is a well-established practice in altered auditory feedback experiments (for a review, see Caudrelier & Rochet-Capellan, 2019), and aims to mask any air-conduction of the participant’s natural (unaltered) voice. This is important, as perception of the unaltered voice could restrict sensorimotor learning with the altered feedback heard through the headphones. When used with clear speech altered feedback, there is clear separation between the two signals; however, it is possible that such masking noise may have resulted in greater masking of the noise vocoded feedback itself, due to greater overlap in the frequency profile of the two signals. We chose to nevertheless include noise masking, for consistency with established conventions in previous literature and to ensure that any failure to observe adaptation could not be attributed to participants’ perception of their unaltered natural voice. It is interesting to note that the study by Borjigin et al., (2024) also used masking noise with both CI and typical hearing participants; thus, the use of masking noise does not seem to completely preclude the observation of significant adaptation with degraded speech. Future studies using auditory perturbations with CI simulated speech could explore the use of alternative maskers with more distinct frequency profiles to noise vocoded speech, such as multi-talker babble.

Overall, the results from this study suggest that successful perceptual learning with CI simulated speech can be seen in the absence of sensorimotor learning when that same CI simulation is applied to real-time speech auditory feedback during production in typical hearing participants. A larger perturbation or longer exposure to CI feedback may be needed to observe sensorimotor learning in this context. This study has further demonstrated that experience of speaking with degraded speech feedback can enhance perceptual learning for understanding that same degraded speech during passive perception, highlighting the potential for engagement in speech production to improve speech perception in CI users who wish to communicate orally. Overall, the use of real-time CI simulated speech auditory feedback offers the potential for more controlled investigation of speech motor control when speaking with degraded auditory feedback, that can have translational implications for speech in CI users.

References

Aubanel, V., & Nguyen, N. (2020). Speaking to a common tune: Between-speaker convergence in voice fundamental frequency in a joint speech production task. *PLOS ONE*, *15*(5). https://doi.org/10.1371/journal.pone.0232209

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. W. (2015). Sensory-motor networks involved in speech production and motor control: An fMRI study. *NeuroImage*, *109*, 418–428. https://doi.org/10.1016/j.neuroimage.2015.01.040

Blamey, P., Sarant, J., Paatsch, L., Barry, J., Bow, C., Wales, R., Wright, M., Psarros, C., Rattigan, K., & Tooher, R. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *JOURNAL OF SPEECH LANGUAGE AND HEARING RESEARCH*, *44*(2), 264–285. https://doi.org/10.1044/1092-4388(2001/022)

Boersma, P. (1993). Accurate Short-Term Analysis Of The Fundamental Frequency And The Harmonics-To-Noise Ratio Of A Sampled Sound. *Proceedings of the Institute of Phonetic Sciences*, *17*.

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer software].

Boisvert, I., Reis, M., Au, A., Cowan, R., & Dowell, R. C. (2020). Cochlear implantation outcomes in adults: A scoping review. *PLOS ONE*, *15*(5), e0232421. https://doi.org/10.1371/journal.pone.0232421

Borjigin, A., Bakst, S., Anderson, K., Litovsky, R. Y., & Niziolek, C. A. (2024). Discrimination and sensorimotor adaptation of self-produced vowels in cochlear implant users. *The Journal of the Acoustical Society of America*, *155*(3), 1895–1908. https://doi.org/10.1121/10.0025063

Bosker, H. R. (2021). Using fuzzy string matching for automated assessment of listener transcripts in speech intelligibility studies. *Behavior Research Methods*, *53*(5), 1945–1953. https://doi.org/10.3758/s13428-021-01542-4

Bradshaw, A. R., Lametti, D. R., Shiller, D. M., Jasmin, K., Huang, R., & McGettigan, C. (2023). Speech motor adaptation during synchronous and metronome-timed speech. *Journal of Experimental Psychology: General*, *152*(12), 3476–3489. https://doi.org/10.1037/xge0001459

Bradshaw, A. R., & McGettigan, C. (2021). Convergence in voice fundamental frequency during synchronous speech. *PLOS ONE*, *16*(10), e0258747. https://doi.org/10.1371/journal.pone.0258747

Bradshaw, A. R., Wheeler, E. D., McGettigan, C., & Lametti, D. R. (2024). Sensorimotor learning during synchronous speech is modulated by the acoustics of the other voice. *Psychonomic Bulletin & Review*, *32*, 306–316. https://doi.org/10.3758/s13423-024-02536-x

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S. (2015). *Audapter* [Computer software]. https://github.com/shanqing-cai/audapter_mex

Cai, S., Boucek, M., Ghosh, S., Guenther, F., & Perkell, JS. (2008). A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. *Proceedings of the 8th Intl. Seminar on Speech Production*, 65–68.

Casserly, E. D. (2015). Effects of real-time cochlear implant simulation on speech production. *The Journal of the Acoustical Society of America*, *137*(5), 2791–2800. https://doi.org/10.1121/1.4916965

Casserly, E. D., & Marino, F. R. (2024). Mirrors and toothaches: Commonplace manipulations of non-auditory feedback availability change perceived speech intelligibility. *Frontiers in Human Neuroscience*, *18*. https://doi.org/10.3389/fnhum.2024.1462922

Casserly, E. D., & Pisoni, D. (2013). Nonword Repetition as a Predictor of Long-Term Speech and Language Skills in Children With Cochlear Implants. *OTOLOGY & NEUROTOLOGY*, *34*(3), 460–470. https://doi.org/10.1097/MAO.0b013e3182868340

Casserly, E. D., Wang, Y., Celestin, N., Talesnick, L., & Pisoni, D. B. (2018). Supra-Segmental Changes in Speech Production as a Result of Spectral Feedback Degradation: Comparison with Lombard Speech. *LANGUAGE AND SPEECH*, *61*(2), 227–245. https://doi.org/10.1177/0023830917713775

Castellanos, A., Benedi, J., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *SPEECH COMMUNICATION*, *20*(1–2), 23–35. https://doi.org/10.1016/S0167-6393(96)00042-8

Caudrelier, T., & Rochet-Capellan, A. (2019). *Changes in speech production in response to formant perturbations: An overview of two decades of research*.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7), 2653–2658. https://doi.org/10.1073/pnas.1216827110

Cowie, R., Douglas-Cowie, E., & Kerr, A. (1982). A study of speech deterioration in post-lingually deafened adults. *Journal of Laryngology and Otology*, *96*(2), 101–112. https://doi.org/10.1017/S002221510009229X

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives; Perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, *25*(4), 1458–1467. https://doi.org/10.3758/s13423-018-1494-x

Franken, M. K., Eisner, F., Acheson, D. J., McQueen, J. M., Hagoort, P., & Schoffelen, J.-M. (2018). Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production. *NeuroImage*, *179*, 326–336. https://doi.org/10.1016/j.neuroimage.2018.06.061

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., & Lind, A. (2021). Speaking With an Alien Voice: Flexible Sense of Agency During Vocal Production. *Journal of Experimental Psychology-Human Perception and Performance*, *47*(4), 479–494. https://doi.org/10.1037/xhp0000799

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., & Lind, A. (2023). Don't blame yourself: Conscious source monitoring modulates feedback control during speech production. *Quarterly Journal of Experimental Psychology*, *76*(1), 15–27. https://doi.org/10.1177/17470218221075632

Gautam, A., Brant, J. A., Ruckenstein, M. J., & Eliades, S. J. (2020). Real-time feedback control of voice in cochlear implant recipients. *Laryngoscope Investigative Otolaryngology*, *5*(6), 1156–1162. https://doi.org/10.1002/lio2.481

Gautam, A., Naples, J. G., & Eliades, S. J. (2019). Control of speech and voice in cochlear implant patients. *The Laryngoscope*, *129*(9), 2158–2163. https://doi.org/10.1002/lary.27787

Gessinger, I., Raveh, E., Steiner, I., & Möbius, B. (2021). Phonetic accommodation to natural and synthetic voices: Behavior of groups and individuals in speech shadowing. *Speech Communication*, *127*, 43–63. https://doi.org/10.1016/j.specom.2020.12.004

Guenther, F. H. (2016). *Neural Control of Speech*. The MIT Press.

Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *The Journal of the Acoustical Society of America*, *123*(1), 273–278. https://doi.org/10.1121/1.2817357

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research*, *130*(2), 133–141. https://doi.org/10.1007/s002219900237

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*(2), 180–190. https://doi.org/10.1111/j.1469-8986.2005.00272.x

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(2), 460–474. https://doi.org/10.1037/0096-1523.34.2.460

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216. https://doi.org/10.1126/science.279.5354.1213

IEEE Subcommittee on Subjective Measurements. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 227–246.

Kim, K. S., Wang, H., & Max, L. (2020). It’s About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *JOURNAL OF STATISTICAL SOFTWARE*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Current Biology*, *28*(19), 3106-3113.e2. https://doi.org/10.1016/j.cub.2018.07.030

Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech and Hearing Research*, *14*(4), 677–709.

Lane, H., & Webster, J. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, *89*(2), 859–866. https://doi.org/10.1121/1.1894647

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862–2872. https://doi.org/10.1121/1.2715657

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, *25*(6), 1198–1205. https://doi.org/10.1177/0956797614529797

Loucks, T. M., Suneel, D., & Aronoff, J. M. (2015). Audio-vocal responses elicited in adult cochlear implant users. *The Journal of the Acoustical Society of America*, *138*(4), EL393–EL398. https://doi.org/10.1121/1.4933233

Max, L., & Daliri, A. (2019). Limited Pre-Speech Auditory Modulation in Individuals Who Stutter: Data and Hypotheses. *Journal of Speech Language and Hearing Research*, *62*(8, S, SI), 3071–3084. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0358

Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience Letters*, *591*, 25–29. https://doi.org/10.1016/j.neulet.2015.02.012

Menard, L., Polak, M., Denny, M., Burton, E., Lane, H., Matthies, M. L., Marrone, N., Perkell, J. S., Tiede, M., & Vick, J. (2007). Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. *JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA*, *121*(6), 3790–3801. https://doi.org/10.1121/1.2710963

Merrikhi, Y., Ebrahimpour, R., & Daliri, A. (2018). Perceptual manifestations of auditory modulation during speech planning. *Experimental Brain Research.*, *236*(7), 1963–1969.

Miller, H., Kearney Elaine, Nieto-Castañón Alfonso, Falsini Riccardo, Abur Defne, Acosta Alexander, Chao Sara-Ching, Dahl Kimberly L., Franken Matthias, Heller Murray Elizabeth S., Mollaei Fatemeh, Niziolek Caroline A., Parrell Benjamin, Perrachione Tyler, Smith Dante J., Stepp Cara E., Tomassi Nicole, & Guenther Frank H. (2023). Do Not Cut Off Your Tail: A Mega-Analysis of Responses to Auditory Perturbation Experiments. *Journal of Speech, Language, and Hearing Research*, *66*(11), 4315–4331. https://doi.org/10.1044/2023_JSLHR-23-00315

Murphy, T. K., Nozari, N., & Holt, L. L. (2023). Transfer of statistical learning from passive speech perception to speech production. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02399-8

Oller, D., & Eilers, R. (1988). The role of audition in infant babbling. *Child Development*, *59*(2), 441–449. https://doi.org/10.1111/j.1467-8624.1988.tb01479.x

Ozker, M., Yu, L., Dugan, P., Doyle, W., Friedman, D., Devinsky, O., & Flinker, A. (2024). Speech-induced suppression and vocal feedback sensitivity in human cortex. *eLife*, *13*, RP94198. https://doi.org/10.7554/eLife.94198

Pardo, J. S., Pellegrino, E., Dellwo, V., & Möbius, B. (2022). Special issue: Vocal accommodation in speech communication. *Journal of Phonetics*, *95*, 101196. https://doi.org/10.1016/j.wocn.2022.101196

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics*, *79*(2), 637–659. https://doi.org/10.3758/s13414-016-1226-0

Parrell, B., & Houde, J. F. (2019). Modeling the Role of Sensory Feedback in Speech Motor Control and Learning. *Journal of Speech Language and Hearing Research*, *62*(8, S, SI), 2963–2985. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA*, *145*(3), 1456–1481. https://doi.org/10.1121/1.5092807

Patel, S., Nishimura, C., Lodhavia, A., Korzyukov, O., Parkinson, A., Robin, D. A., & Larson, C. R. (2014). Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback. *The Journal of the Acoustical Society of America*, *135*(5), 3036–3044. https://doi.org/10.1121/1.4870490

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2217714

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, *106*(6), 3629–3636. https://doi.org/10.1121/1.428215

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Shiller, D. M., Bobbitt, S., & Lametti, D. R. (2023). Immediate cross-language transfer of novel articulatory plans in bilingual speech. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001456

Shiller, D. M., Mitsuya, T., & Max, L. (2020). Exposure to Auditory Feedback Delay while Speaking Induces Perceptual Habituation but does not Mitigate the Disruptive Effect of Delay on Speech Auditory-motor Learning. *NEUROSCIENCE*, *446*, 213–224. https://doi.org/10.1016/j.neuroscience.2020.07.041

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *BRAIN AND LANGUAGE*, *164*, 77–105. https://doi.org/10.1016/j.bandl.2016.10.004

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research*. https://doi.org/10.1044/jshr.1804.795

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-Down Influences of Written Text on Perceived Clarity of Degraded Speech. *Journal of Experimental Psychology- Human Perception and Performance*, *40*(1), 186–199. https://doi.org/10.1037/a0033206

Svirsky, M. A., Silveira, A., Surez, H., Neuburger, H., Lai, T. T., & Simmons, P. M. (2001). Auditory Learning and Adaptation after Cochlear Implantation: A Preliminary Study of Discrimination and Labeling of Vowel Sounds by Cochlear Implant Users. *Acta Oto-Laryngologica*, *121*(2), 262–265. https://doi.org/10.1080/000164801300043767

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3). https://doi.org/10.1016/J.NEUROIMAGE.2007.09.054

Tyler, R., & Summerfield, A. (1996). Cochlear implantation: Relationships with research on auditory deprivation and acclimatization. *EAR AND HEARING*, *17*(3, S), S38–S50. https://doi.org/10.1097/00003446-199617031-00005

Van Summers, W., Pisoni, D., Bernacki, R., Pedlow, R., & Stokes, M. (1988). Effects of noise on speech production- acoustic and percpetual analyses. *Journal of the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2773966

Zheng, Z. Z., MacDonald, E. N., Munhall, K. G., & Johnsrude, I. S. (2011). Perceiving a Stranger’s Voice as Being One’s Own: A ‘Rubber Voice’ Illusion? *PLOS ONE*, *6*(4), e18655.