Experimental Brain Research
https://doi.org/10.1007/s00221-018-5248-9

RESEARCH ARTICLE

1839

The impact of perilaryngeal vibration on the self-perception of loudness and the Lombard effect

François-Xavier Brajot[1,2,3] · Don Nguyen[2,3] · Jeffrey DiGiovanni[1] · Vincent L. Gracco[2,3,4]

Received: 25 September 2017 / Accepted: 29 March 2018

Abstract
The role of somatosensory feedback in speech and the perception of loudness was assessed in adults without speech or hearing disorders. Participants completed two tasks: loudness magnitude estimation of a short vowel and oral reading of a standard passage. Both tasks were carried out in each of three conditions: no-masking, auditory masking alone, and mixed auditory masking plus vibration of the perilaryngeal area. A Lombard effect was elicited in both masking conditions: speakers unconsciously increased vocal intensity. Perilaryngeal vibration further increased vocal intensity above what was observed for auditory masking alone. Both masking conditions affected fundamental frequency and the first formant frequency as well, but only vibration was associated with a significant change in the second formant frequency. An additional analysis of pure-tone thresholds found no difference in auditory thresholds between masking conditions. Taken together, these findings indicate that perilaryngeal vibration effectively masked somatosensory feedback, resulting in an enhanced Lombard effect (increased vocal intensity) that did not alter speakers' self-perception of loudness. This implies that the Lombard effect results from a general sensorimotor process, rather than from a specific audio-vocal mechanism, and that the conscious self-monitoring of speech intensity is not directly based on either auditory or somatosensory feedback.

Keywords Vibration · Speech · Masking · Noise · Lombard effect · Somatosensory feedback

Introduction

The Lombard effect is defined as an increase in vocal intensity due to a concomitant increase in environmental noise (Lombard 1911). Other acoustic properties of speech affected by external noise include fundamental frequency, duration, spectral contour, formant frequencies and bandwidths, and rate of speech (Lee et al. 2007; Junqua 1993; Letowski et al. 1993; Schulman 1989). Perceptual characteristics are likewise modified, including perception of pitch, vocal effort and intelligibility (Junqua 1996; Van Summers et al. 1988; Pickett 1956). The extent of change across these various parameters has been shown to depend on the type of noise presented (stochastic vs. multi-taker babble; Junqua 1993) and the communicative intent of the speaker (Lane and Tranel 1971). The effect is present not only in humans, but across bird and mammal species (Tressler et al. 2011; Scheifele et al. 2005; Cynx et al. 1998). It is a robust neurophysiological response that can be evoked through stimulation of specific brainstem regions (Nonaka et al. 1997) and is difficult to consciously inhibit (Pick et al. 1989; Thierren et al. 2012). It is, in other words, a phylogenetically old mechanism that alters the control and perception of self-produced vocalizations, typically improving intelligibility, with acoustic changes that are similar, but not identical, to loud speech (Letowski et al. 1993; Junqua 1996).

An early explanation and important theoretical paradigm for later models of speech sensorimotor control characterizes the Lombard effect as the product of a regulating servo-mechanism that compensates for disturbances in feedback

✉ François-Xavier Brajot
brajot@ohio.edu

1 Department of Communication Sciences and Disorders, Ohio University, Grover Center W221, Athens, OH 45701, USA

2 Centre for Research on Brain, Language and Music, McGill University, 3640 de la Montagne, Montréal, Québec H3G 2A8, Canada

3 School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montréal, Québec H3G 1A8, Canada

4 Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, USA

Published online: 05 April 2018

(Fairbanks 1954; Kerrison 1918). Very generally, sensory feedback is compared against a desired level and corrections are made for upcoming productions. In the case of the Lombard effect, the desired level is the volume above some reference intensity. Increasing noise reduces that difference, which is then corrected for by speaking more loudly. Within this framework, the Lombard effect is one of a class of responses to auditory feedback manipulations that include compensations for pitch and formant shifting. To contend with the findings that corrections for such feedback perturbations are only incomplete, that is, a fraction of the size of the perturbation, certain authors speculate that the overall effect is the combined effect of both auditory and somatosensory feedback (Katseff et al. 2011; Larson et al. 2008). A compensatory response is preceded by multisensory integration, limiting the impact of perturbation to a single modality. This has been extended to explain differential effects when somatosensory feedback is degraded, either empirically or due to disease (Mu et al. 2013; Liu et al. 2012; Hammer and Barlow 2010; Larson et al. 2008). By reducing the amount or quality of somatosensory feedback, a speaker becomes more dependent on auditory feedback, and, therefore, compensates to a greater degree for perturbations in auditory feedback (Lametti et al. 2012).

A competing hypothesis, raised by Lane and Tranel (1971), denies this parallel between the two forms of sensory feedback, proposing instead that the Lombard effect is the result of a tracking mechanism whose goal is to maintain communicative effectiveness. It is a listener-directed, not speaker-directed, phenomenon. According to these authors, a low-level audio-vocal hypothesis is not consistent with the finding that the magnitude of the Lombard effect is stronger in conversational speech (when a listener is present) than during oral reading or other speech tasks that do not involve some communicative interaction (Garnier et al. 2010). They also point out that sidetone compensation, and specifically a decrease in vocal intensity when a speaker's auditory feedback intensity is increased, does not follow from the premise that vocal intensity changes to "better hear onself". The fundamental component of this argument, however, is based on findings from psychophysical studies that show that altering auditory feedback does not affect a speaker's sense of loudness (Lane et al. 1961).

Loudness is the psychological percept of intensity. It is often measured along some subjective scale relative to physical intensity of a signal. When speakers are asked to rate the loudness of their own speech, as opposed to a recording of the same speech, the relationship of perceived loudness to signal intensity increases along a steeper scale, that is, there is a shift in the *slope* of the loudness curve when comparing passive-listening (sone scale) to self-generated (autophonic scale) speech (Fig. 1, panel 1). This indicates that the act of

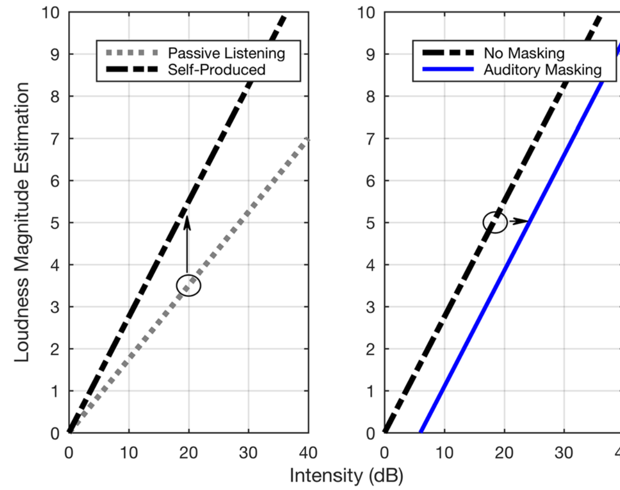

Fig. 1 Examples of typical changes to loudness functions (adapted from Lane et al. 1961, Fig. 6). The slope of the loudness function increases when comparing passive-listening to self-produced speech (left panel). The intercept of the loudness function shifts to the right when a Lombard effect is elicited (right panel). Arrows indicate the direction of change

vocalizing, and the involvement of somatosensory feedback, alters the relative sense of loudness.

Now, with respect to self-generated sounds, speakers compensate for changes in auditory feedback by increasing or decreasing vocal intensity. This corresponds to a shift along the x axis in Fig. 1. The relation of loudness estimates relative to intensity does not change, however. The slope remains invariant. With the Lombard effect, for example, speaking in noise may make someone speak 6 dB louder, but their perceptual rating stays the same (Fig. 1, panel 2). The Lombard effect is, therefore, both an increase in vocal intensity and a lack of awareness of the change in voice.

In short, there is a perceptual shift in loudness when speech is self-generated as opposed to passively listened to. When the speech is self-generated, however, an externally induced change in vocal intensity (Lombard effect) does not affect perceived loudness. The change in vocal intensity does seem to vary as a function of communicative context, however. This suggests that the references, or the information relevant in somatosensory and auditory feedback, are different. Somatosensory feedback functions to monitor one's speech (interoceptive), whereas auditory feedback is referenced to communicative goals (exteroceptive).

Both theoretical approaches involve some form of correction based on feedback. The principal distinction is in the presumed roles for auditory and somatosensory feedback. While many feedback models posit a correction based on the (often additive) combination of both types of feedback, Lane and Tranel (1971) argue that each type provides qualitatively different information with distinct functions in the control of speech. Self-monitoring is still recognized in this latter approach, but is limited to somatosensory feedback.

The current study tests these two approaches by assessing the effect of combined auditory-somatosensory masking on speech. Specifically, we assessed the effect of adding external vibration to the neck on the Lombard effect and perceptual ratings of loudness. A change in the slope of the loudness curve would support the hypothesis that somatosensory feedback is relevant for perceived changes in intensity. Invariance of the slope of the curve with a shift in intercept would support the hypothesis that somatosensory feedback plays a complementary role in eliciting the Lombard effect, requiring an alternative explanation for loudness perception. A shift in intercept should furthermore be associated with an overall shift of the intensity contour in running speech. Further analyses were carried out on additional acoustic variables (fundamental frequency and formant frequencies), as well as auditory thresholds, to confirm that the effects of vibration were indeed impacting somatosensory feedback.

Methods

Ethics statement

This study was conducted in accordance with the 1964 Helsinki declaration and its later amendments. Informed consent, approved by the McGill School of Medicine Institutional Review Board, was reviewed and signed by all participants included in this study.

Subjects

25 adults were recruited for this study. Two subjects were not able to complete the experiment due to equipment malfunction. One subject's data were thrown out due to medical history. The remaining 22 subjects consisted of 13 females and 9 males; 18–59 years of age (mean 28.3, SD 10.1). They reported no history of neurological, psychiatric, speech or hearing disorders. All subjects passed a hearing screening with a minimum threshold of 20 dB HL tested at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz.

Instrumentation

Speech was recorded with a headset microphone (AKG C420), positioned 4–5 cm from the subject's mouth, onto a digital audio recorder (M-Audio MicroTrack 24/96) set at 16-bit resolution and 44.1 kHz sampling rate. All auditory signals were presented using Etymotic ER-2 insert earphones via an Aphex headpod amplifier. Earphones were inserted deeply, following manufacturer recommendations, to reach the second bend of the ear canal and minimize occlusion effects (Dean and Martin 2000). Signal intensity in dB SPL was pre-calibrated using a digital sound level meter (Radio Shack 33-2055; A-weighting) and 2 cc coupler. The vibratory signal was presented using a Brüel and Kjaer minishaker (Model 4810), driven by a Brüel and Kjaer amplifier (Model 2716C). The minishaker is a lightweight, electrodynamic exciter well suited for speech research as it allows precise delivery of low force vibration to small object (Andreatta and Barlow 2009). An Endevco 2311-500 miniature force transducer attached to the minishaker acted as the point of contact to the subject's skin, placed at the level of the left thyroid lamina. Transducer contact was 12.7 mm in diameter, applied with double-sided adhesive washers commonly used for surface electrode placement.

Stimuli

The auditory masking signal was a pink noise with a 20 dB/decade roll-off between 1000 and 4500 Hz and cut-off frequencies at 200 and 4500 Hz. This not only effectively masked frequencies important to human speech, but also reduced listener discomfort at higher intensities. The vibrotactile masking signal input to the minishaker was a low-pass filtered version (cut-off at 500 Hz and a 50 dB/decade roll-off) of the speech-weighted noise used in the auditory masking condition. These lower frequencies were selected to mask the typical range of vocal fundamental frequencies while reducing the damping effect of higher frequencies on the minishaker displacement amplitudes. All stimuli were generated and output from custom programs written in Matlab.

Procedures

Loudness magnitude estimation

Participants were prompted to say the sound /ʌ/ at varying volumes, following an instruction to say the sound either "normally" or "more loudly" or "more softly" than the immediately preceding production. After each utterance, participants were asked to rate the loudness of their production using any arbitrary scale of their choice. This is an "absolute magnitude estimation" method (Zwislocki and Goodman 1980) that does not impose a pre-assigned modulus (a number corresponding to the center of the speaker's range assigned by the experimenter) to be able to identify differences in loudness function intercepts. To familiarize themselves with the task, all subjects first completed a practice run that had them move up and down their selected intensity scales. The experimental phase consisted of three conditions, each composed of 40 repetitions. There were 10 "normal", 15 "louder" and 15 "softer" instructions, presented in random order in each condition. This design was implemented to avoid over-sampling "normal" productions and increase the likelihood that subjects produce volumes beyond any initially anticipated range limits. The three conditions were always presented in the same order: repetition of /ʌ/ under (1) normal feedback, (2) auditory masking, and (3) mixed auditory plus vibrotactile masking. The auditory speech-weighted masking signal was presented at 90 dB bilaterally in both auditory and mixed masking conditions.

Lombard effect

Participants read the first few sentences of the Rainbow Passage (Fairbanks 1960) aloud. Three readings were completed under three separate masking conditions: (1) a baseline (reference) condition with no extraneous auditory input, (2) an auditory masking condition with speech-weighted noise presented binaurally through insert earphones at 90dB SPL, and (3) a mixed (auditory plus laryngeal vibrotactile) masking condition with combined speech-weighted noise presented at the ear as well as a low-pass filtered version presented at the neck with the minishaker. To verify that a Lombard effect was elicited within speakers, masking was applied 400 ms after the prompt to vocalize and continued for 1 s. Masking onset typically occurred on the fourth or fifth syllable of each production, depending on when the subject began vocalizing.

Some of the energy from vibration at the neck is conducted to the ear, resulting in an increase in auditory threshold and a concomitant Lombard effect when auditory feedback is not masked (see post hoc analysis). In such a case, it is impossible to separate somatosensory from auditory effects. Because of this, a vibration-only condition was considered not to be a proper control and was not included in the experimental procedure.

Data analysis

Loudness magnitude estimation

Speech signals were extracted from recordings using a semi-automatic segmentation routine that marked 15% of the peak amplitude on each side. Root mean square amplitude (RMS) was calculated for each production then converted to dB SPL. RMS was calculated by applying a Hamming window to each consecutive 20 ms of the signal, followed by zero-phase filtering (Matlab filtfilt) and taking the square root of the absolute value of the filtered output. Correlation coefficients were calculated on the raw data, after removing missing values (tokens where subjects failed to vocalize or were otherwise clipped). Slope and intercept of linear fits to the data in each condition were calculated using the Matlab polyfit function. Fitted lines were normalized to a 0–10 magnitude estimation scale. All conditions were then referenced to the normal feedback condition so that altered feedback condition intercepts represent baseline dB difference relative to normal feedback. Note that since intensity values are reported in decibels, the difference between masked and baseline productions is equal to describing signal gain (the log of masked signal power to baseline signal power).

Lombard effect

Recorded signals were down-sampled to 8820 Hz and RMS envelopes were calculated for the entire 10 s of the recording. Peaks in the RMS envelopes consistent across productions were identified using a custom program, resulting in the following 35 syllabic groups for which the RMS peak values were retained:

When the sun | light | strikes | rain | drops | in the | air,
| they | act | like a | prism | and | form a | rain | bow.
The rain | bow is | a | di | vi | sion of | white | light | into
| many | beauti | ful | colors.
These | take | the | shape | of a | long | round (…)

Selected RMS values were then converted to dB SPL. Signal fundamental frequency (f_0), and first and second formant values (F1 and F2) were also calculated for the syllables marked in the intensity analysis.

Intensity and frequency values for marked syllables were averaged within subjects, then separated into one of three masking segments: before, during or after masking. A multivariate analysis of variance was completed to compare dependent variables as a function of masking—similar to the analysis of magnitude estimates. Intensity and frequency measures were averaged across syllables within a masking segment. A principal component analysis was subsequently carried out to control for multicollinearity among acoustic variables. Analyses of variance were then carried out on the principal components. To simplify the analysis, pre- and post-masking segments were not included as they did not provide new information in the latter context. Statistical analyses were carried out with the Statistica software package.

As perilaryngeal vibration is a fairly novel method for masking somatosensory feedback of the voice (cf. Loucks and DeNil 2012), two methods were employed to establish validity. The first was to extend the acoustic analysis to other acoustic variables affected by a Lombard effect: fundamental frequency (f_0), first ($F1$) and second ($F2$) formant frequencies (Junqua 1993). This allowed us to more fully characterize the effects of vibration and identify differences between auditory and mixed masking conditions. The second was to measure pure-tone thresholds in each of the experimental conditions. A lack of difference between auditory and mixed masking conditions, in this case, would support the claim that vibration did impact somatosensory feedback and was not the result of an auditory effect due to bone-conduction and occlusion effect. Thresholds in a worst-case scenario (no-masking and over-the-ear headphones) were also included for comparison.

Age and sex differences have been shown for all speech acoustic variables included in this study (Baken and Orlikoff 2000). With respect to the Lombard effect, Junqua (1993) found a larger increase in pitch for males and a larger increase in $F1$ for females. $F2$ showed a similar, although non-significant, trend in female speakers. Age and sex were, therefore, included as between subject variables to determine possible influence on experimental outcomes.

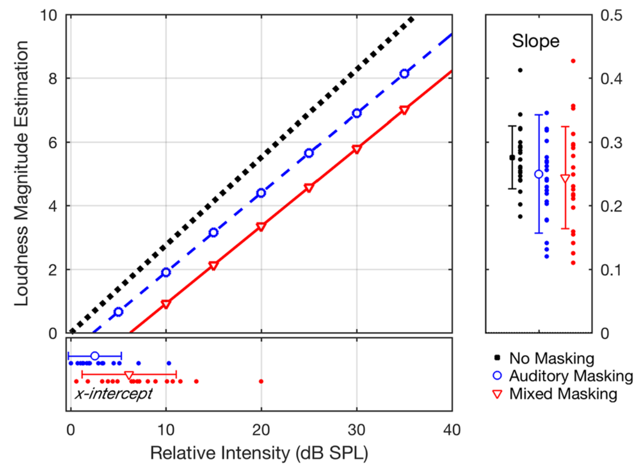

Fig. 2 Mean loudness magnitude estimation functions across conditions. Main panel: group mean loudness curves for baseline (dotted black square), auditory (dashed blue circle) and mixed masking (solid red triangle) conditions. Lower panel: Mean and standard deviation of normalized x-intercepts. Right panel: Mean and standard deviation of slopes in each condition. Individual subject data points are plotted alongside each corresponding error bar

Results

Loudness magnitude estimation

All subjects reported greater difficulty attributing loudness magnitude estimates in masking conditions, and mixed masking was generally felt to be more difficult than auditory masking alone. Nevertheless, correlation coefficients between speech intensity and loudness magnitude estimates remained high across conditions: $r = 0.88$ ($p < 0.01$, 95% confidence interval 0.74–0.95) for normal feedback, $r = 0.89$ ($p < 0.01$, 95% confidence interval 0.76–0.95) for auditory masking, and $r = 0.87$ ($p < 0.01$, 95% confidence interval 0.73–0.94) for mixed masking. A Friedman's ANOVA (Friedman 1937) comparing correlation coefficients failed to reach significance, $\chi^2(2) = 1.09$, $p = 0.6$. The presence or type of masking, therefore, did not appear to affect response accuracy (Fig. 2).

A repeated-measures ANOVA on slope failed to show a significant difference across conditions [$F(2,42) = 1.92$, $p = 0.16$]. The relative autophonic loudness scale, therefore, remains invariant across conditions, despite masking of auditory and vibrotactile feedback.

A repeated-measures ANOVA on intercept was significant, [$F(2,42) = 11.6$, $p < 0.001$]. Bonferroni–Holm-corrected multiple comparisons (Holm 1979) showed differences across all conditions, with a greater shift for the mixed masking than the auditory masking condition: normal feedback vs. auditory masking $t(42) = 3.1$, $p = 0.01$, auditory vs. mixed masking $t(42) = 2.2$, $p = 0.03$, normal feedback vs. mixed masking $t(42) = 4.0$, $p < 0.001$. The mean intensity shift under auditory masking was approximately 2.5 dB SPL. The mean intensity shift under mixed masking increased to 6 dB SPL.

Lombard effect

A multivariate analysis was completed using the general linear model subroutine in Statistica. Condition (none, auditory, mixed) and section (before, during, and after masking) were specified as within-subjects factors, sex as a categorical predictor and age as continuous predictor. Using Wilks' statistic, no section x condition interactions were found: section × condition × sex $\Lambda = 0.29$, $F(16,4) = 0.61$, $p = 0.79$, section × condition × age $\Lambda = 0.18$, $F(16,4) = 1.12$, $p = 0.51$, section × condition $\Lambda = 0.07$, $F(16,4) = 3.51$, $p = 0.12$. Age was not a significant predictor: condition × age $\Lambda = 0.789$, $F(8, 12) = 0.4$, $p = 0.90$, section × age $\Lambda = 0.395$, $F(8, 12) = 2.3$, $p = 0.10$, age $\Lambda = 0.949$, $F(4, 16) = 0.21$, $p = 0.93$. Sex was a significant predictor: condition × sex $\Lambda = 0.164$, $F(8, 12) = 7.67$, $p = 0.001$, section × sex $\Lambda = 0.197$, $F(8, 12) = 6.11$, $p = 0.003$, sex $\Lambda = 0.122$, $F(4, 16) = 28.92$, $p < 0.001$. Main effects of condition, $\Lambda = 0.260$, $F(8, 12) = 4.3$, $p = 0.012$, and section, $\Lambda = 0.047$, $F(8, 12) = 30.61$, $p < 0.001$, were also identified.

These results may be considered with respect to Fig. 3, which plots section-by-condition means with 95% confidence intervals for each of the four measured dependent variables, separated by sex. Values for all variables are similar before masking onset, shift away from baseline during masking, then return to baseline levels after masking (effect of masking section). Adding vibration to the neck increases effects observed under auditory masking only (effect of condition). Compared to male subjects, female subjects show expected higher fundamental frequencies and slightly lower intensities (interaction with sex). $F1$ displacement due to

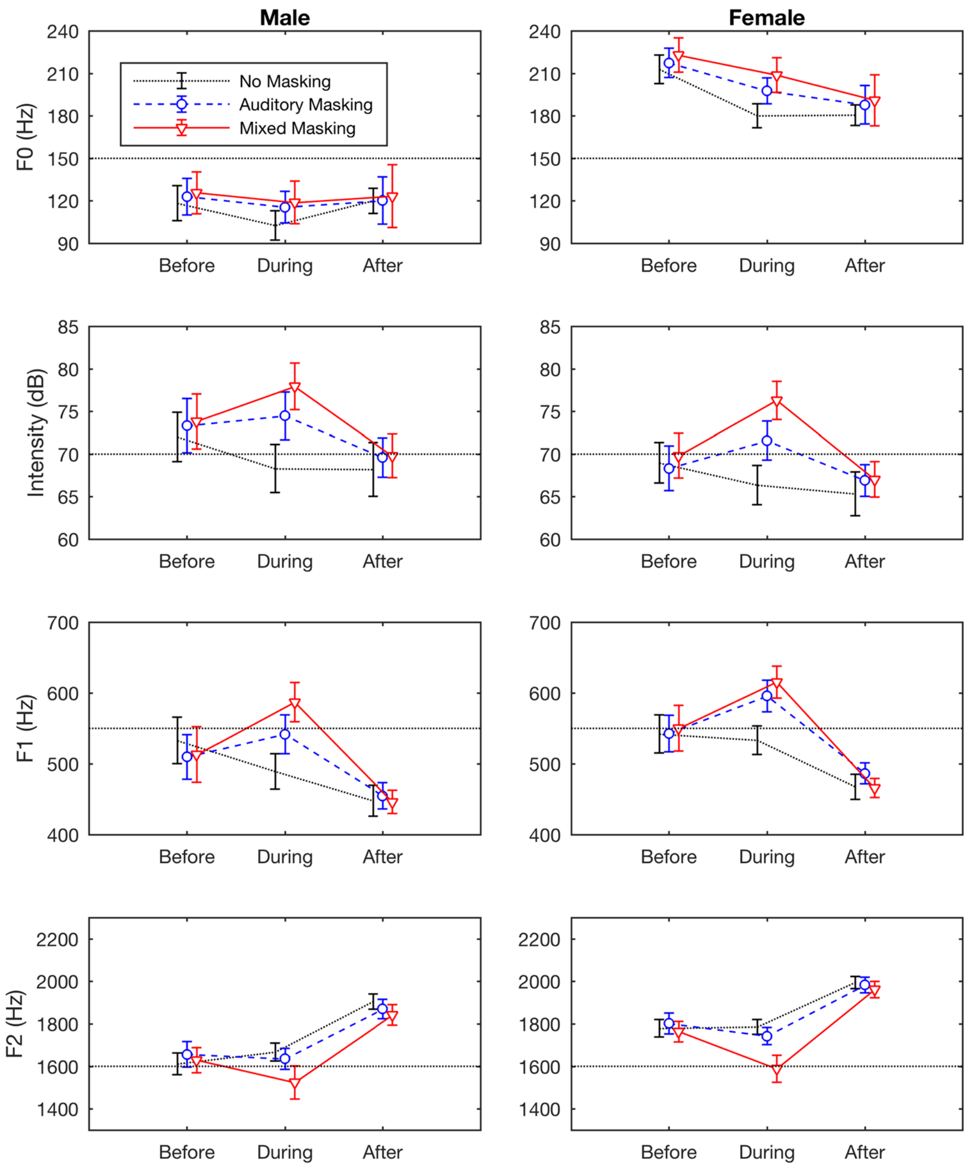

Fig. 3 Mean and 95% confidence intervals for dependent variables before, during, and after masking. The baseline condition is plotted dotted black, auditory masking dashed blue with circles, and mixed masking solid red with triangles

masking in the auditory condition is greater for female than male subjects, consistent with previous work (Junqua 1993), resulting in greater overlap between masking conditions. $F2$ is more clearly affected by the addition of vibrotactile masking than auditory masking only.

Since no condition-by-section interaction was found, and because we are primarily interested in effects of masking, we focused subsequent analyses on the "during masking" section alone. Age was not included as a predictor either, as it was found to be non-significant in the omnibus test. A multivariate analysis with condition as within-subjects factor and sex as categorical predictor revealed a significant condition × sex interaction, $\Lambda = 0.339$, $F(8,13) = 3.17$, $p = 0.03$, and main effects of condition, $\Lambda = 0.037$, $F(8,13) = 42.75$, $p < 0.001$, and sex, $\Lambda = 0.097$, $F(4,17) = 39.68$, $p < 0.001$.

A follow-up discriminant analysis on this data subset was not appropriate due to multicollinearity across acoustic variables (continuous predictors in the discriminant analysis); correlations between variables were as high as 0.7 in certain cases. This is to be expected, given the known interrelationships among speech acoustic parameters, such as covariation between f_0 and dB (Titze 1992). To control for this, we conducted a principal component analysis on the four variables, with oblique rotation. This reduced the variables to three components representing the factors of sex, auditory masking and vibration.

Factor loadings for the resulting principal components are provided in Table 1. The first principal component accounted for 36% of the total variance, and is associated with important acoustic indices of speech intensity: dB and $F1$. The second component accounted for 30% of the

Table 1 PCA standardized loadings (pattern matrix) after oblique translation

	1. Auditory	2. Sex	3. Vibratory	Communality (h^2)	Uniqueness (u^2)
f_0	0.00	**0.99**	– 0.03	0.96	0.039
dB	**0.88**	– 0.24	– 0.20	0.92	0.084
F1	**0.80**	**0.38**	0.14	0.90	0.096
F2	– 0.02	– 0.05	**1.01**	0.99	0.011

Bold text denotes loadings greater than 0.3

total variance and is more clearly linked with sex-related acoustics—namely f_0, but also with contributions from $F1$ and dB. The third component accounted for 28% of the total variance and is largely dependent on $F2$ changes associated with adding vibrotactile masking; smaller contributions from dB and $F1$ are also present.

Data are plotted as a function of the two components that relate to experimental conditions in Fig. 4. The histograms along each axis help to represent the differences in distributions along a single dimension. The first, "Lombard" component distinguishes among conditions, with greater overlap between masking conditions. The third, "Vibration" component clearly separates the mixed masking from the other conditions. Components relevant to sex effects are plotted in Fig. 5. A clear distinction can be seen along the "Sex" component associated with f_0 variation. The "Vibration" component also shows a degree of sex-specific variation, however, both in size and center of distribution. Note that differences in histogram amplitudes are due to unequal sample sizes as a function of sex.

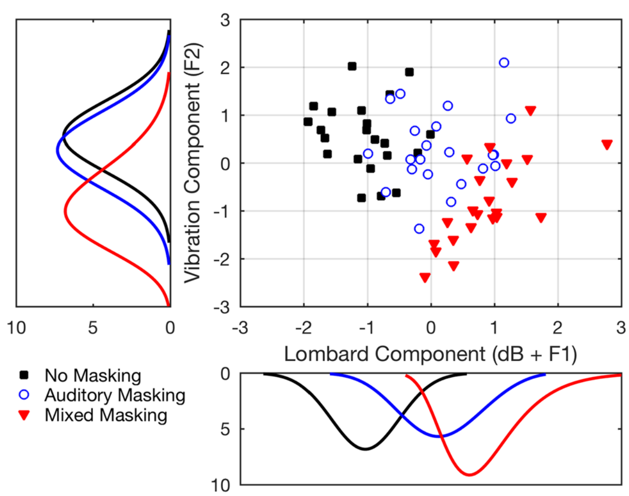

Fig. 4 Projection of cases on "Lombard" vs. "Vibration" factor plane. Unmasked, auditory and mixed masking conditions are plotted with black squares, blue circles and red triangles, respectively. Distributions along each axis represent approximate frequency distributions

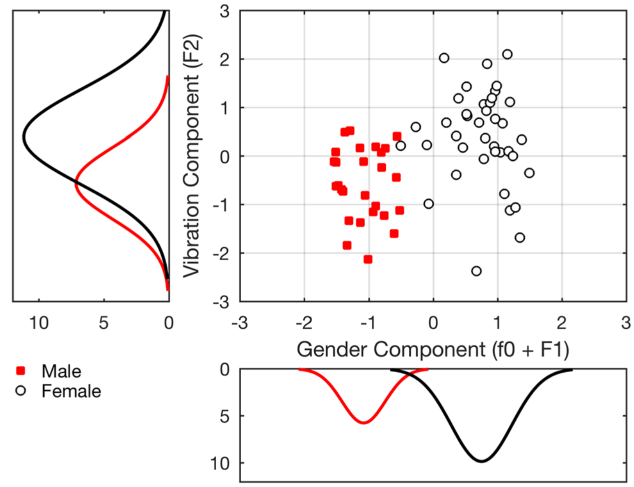

Fig. 5 Projection of cases on "Sex" vs. "Vibration" factor plane. Data for female subjects are plotted with open black circles, those for male subjects with filled red squares. Distributions along each axis represent approximate frequency distributions (Gaussian fit to histogram); amplitude differences are due to unequal group size

Factor scores from the principal component analysis were subsequently submitted to a multivariate analysis, with condition as within-subjects factor and sex as a categorical predictor. No condition × sex interaction was found, $F(6,15)=2.21$, $p=0.1$, but there were main effects of condition, $F(6,15)=59.47$, $p<0.001$, and sex, $F(3,18)=50.25$, $p<0.001$. Follow-up univariate ANOVAs were conducted on each of the components, with sex as categorical predictor. Results are displayed in Table 2. No interactions effects were found, but main effects of condition were present across components. Not surprisingly, significant main effects of sex were also identified for components 2 and 3, in line with sex-related differences in f_0 and $F2$.

Bonferroni–Holm-corrected post hoc comparisons of conditions for the "Lombard" component confirmed significant difference across conditions: baseline vs. auditory $t(42)=-6.63$, $p<0.001$, auditory vs. mixed $t(42)=-3.66$, $p=0.006$, baseline vs. mixed $t(42)=-10.63$, $p<0.001$. Bonferroni–Holm-corrected post hoc comparisons of conditions for the "Vibration" component revealed no significant difference among male subjects for baseline vs. auditory, $t(42)=1.32$, $p=0.82$, marginal significance for auditory vs. mixed, $t(42)=2.93$, $p=0.058$, and a significant difference for baseline vs. mixed, $t(42)=3.98$, $p=0.01$. Among female subjects, baseline vs. auditory was also non-significant, $t(42)=1.16$, $p=0.45$, but the other comparisons reached significance: auditory vs. mixed, $t(42)=4.07$, $p=0.005$, and baseline vs. mixed, $t(42)=5.21$, $p<0.001$. Post hoc comparison of the "Sex" component revealed a significant difference for female subjects comparing baseline and mixed

Table 2 Results of univariate analyses on individual components

Effects	DOF	Lombard (dB + $F1$)		Sex (f_0 + $F1$)		Vibration ($F2$)	
		F	p	F	p	F	p
Sex	1, 20	0.10	0.81	**154.6**	**< 0.001***	**14.44**	**0.001***
Condition	2, 40	**114.3**	**< 0.001***	**61.4**	**< 0.001***	**66.14**	**< 0.001***
Condition × sex	2, 40	0	0.96	3.10	0.06	1.86	0.17

Bold font and asterisks denote $p < 0.05$

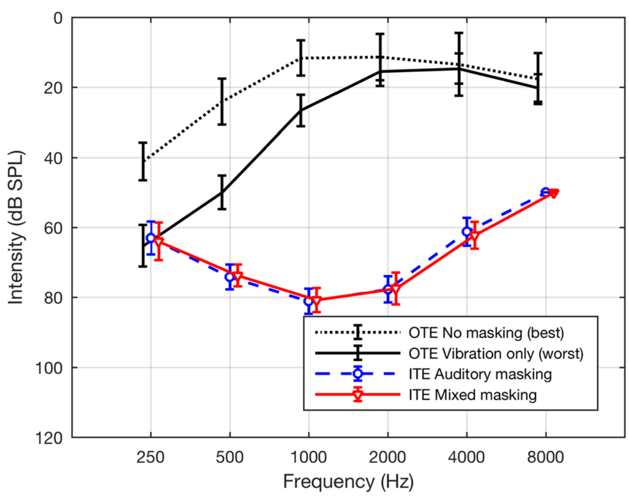

Fig. 6 Pure-tone hearing thresholds comparing effects of vibration and auditory masking

masking conditions only, $t(42) = -3.71$, $p = 0.01$. All other comparisons were non-significant.

Post hoc perceptual analysis

To further ascertain that the observed effects were indeed due to perilaryngeal vibration and not additional auditory stimulation from non-osseous sound conduction and an occlusion effect (Adelman et al. 2012), together or in isolation, we conducted an additional pure-tone hearing threshold assessment in each of the experimental conditions. 18 adults were recruited for this analysis (mean age of 28 years, 7.7.2 SD; 6 male, 12 female). Informed consent, approved by the McGill School of Medicine Institutional Review Board, was reviewed and signed by all participants included in this study. Pure tones were presented at 250, 500, 1000, 2000, 4000 and 8000 Hz.

We first tested subjects with a standard audiometer using over-the-ear headphones to assess the impact of vibration on baseline hearing thresholds in a "worst-case" scenario, that is, with known occlusion effect (ear canal occluded at the outer ear). The dashed and solid black lines in Fig. 6 represent group thresholds for the over-the-ear no-masking and vibration-only conditions, respectively. Perilaryngeal vibration visibly increases auditory thresholds for frequencies below 2000 Hz. Non-osseous sound conduction and occlusion effects are, therefore, clearly relevant confounds for conditions with perilaryngeal vibration. We then tested subjects with the experimental set-up (with insert earphones) across masking conditions, including vibration-only. Thresholds for no-masking and vibration-only fell below the minimum calibration level of 50 dB SPL. Results for the auditory only (blue circles) and mixed (red triangles) masking conditions are plotted in Fig. 6 for comparison against "worst-case" over-the-ear vibration condition. The results show that perilaryngeal vibration did not affect thresholds beyond levels in the auditory masking condition alone, supporting the conclusion that effects of vibration in the mixed condition cannot be simply attributed to an additional auditory effect.

Discussion

The role of somatosensory feedback in the control of speech intensity was evaluated by masking perilaryngeal vibrotactile feedback during speech tasks. Perilaryngeal vibration effectively increased vocal intensity beyond what was observed for auditory masking alone, indicating that both auditory and somatosensory feedback contribute to eliciting the Lombard effect. Similarly, the shift in loudness estimates attributed to the Lombard effect was greater when auditory masking was paired with vibratory masking. The slope of the loudness function did not change, however, indicating that speakers were unaware of the added effect of vibration on their speech. Overall, these findings suggest that somatosensory feedback is part of a general sensorimotor mechanism that adjusts speech intensity without directly impacting the self-perception of loudness.

A generalization of the definition of the Lombard effect as a response to noise in either auditory or somatosensory feedback is in line with interpretations of other compensatory responses. Unaltered somatosensory feedback has been implicated in partial adaptation to a variety of auditory feedback perturbations (Katseff et al. 2011). Reduced somatosensory feedback has also been associated with larger compensations to shifts in vocal pitch (Larson et al. 2008). Results of the current experiment are simply an extension of this pattern to adjustments in vocal intensity as noise is introduced to both auditory and somatosensory systems.

Generalizing the effect across sensory systems does not require positing a single or uniform mechanism, however. Recent models that have elaborated on the servomechanism analogy distinguish contributions from separate feedback systems that nevertheless work toward the same overall goal. Expanding on the work by Houde and colleagues (Houde and Nagarajan 2011), for example, Hickok (2012) proposes a hierarchical feedback model that pairs auditory and somatosensory feedback with lexical and phonetic processes, respectively. Each type of feedback is associated with internal predictions framed by different organizational principles, but that necessarily interact along the forward path (output) of the hierarchy. This allows for variability in the presentation of a response, such as increased duration for vowels (Junqua 1993) or the drop in $F2$ under perilaryngeal vibration observed in the current study. Auditory feedback is more directly influenced by higher-level linguistic requirements within the hierarchy, moreover, which may help explain a Lombard effect susceptive to communicative context (Amazi and Garber 1982; Lane and Tranel 1971). Mutual interaction between sensory modalities allows the system to be less sensitive to unisensory perturbation (i.e., partial compensation). The lexical-phonetic hierarchy furthermore relates easily to the exteroceptive–interoceptive distinction proposed by Lane and Tranel (1971).

The question of the conscious perception of self-produced speech is not frequently addressed in sensorimotor models of speech production. This may be because the automatic nature of the compensatory responses that are so often the basis of their design presupposes that volitional control lies outside the presumed mechanisms. The findings from the current study certainly support this premise. As was noted in the introduction, Lane and Tranel (1971) argued that invariance in the slope of the loudness function across listening conditions suggests that auditory feedback does not play a primary role in loudness perception nor in "direct feedback control during speech" (p. 694). Similar slopes between auditory and mixed masking conditions in the current study extend this argument to somatosensory feedback. This implies that the self-perception of loudness is not based on sensory feedback and must be referenced elsewhere.

The implication for models of sensorimotor control is that a separate loop or set of components must be included to account for volitional responses and perception. This of course complicates the model, but is not a new notion nor inconsistent with current neurobiology. Over six decades ago, Lee (1950) had already posited that higher-level word and thought loops in speech production involved conscious processes that were updated by lower sensorimotor corrective loops only if the production was correct (cf. Hickok 2012; Fautrelle and Bonnetblanc 2012). Guenther et al. (2006) returned to this notion in specifying that predicted sensory consequences from their model's speech sound map are "based on prior successful attempts" (p. 286). Self-perception is thus internally referenced and dependent on *correct* productions over multiple iterations, and by extension over a broader time scale than that usually referenced for compensatory responses.

With the emphasis on forward models in recent years (Houde and Nagarajan 2011; Hickok 2012; Tian and Poeppel 2012), feedforward representations have become an obvious candidate for the perception of self-generated sounds and provide potential explanations for perceptual deficits specific to self-generated speech (Brajot et al. 2016). Work on mental imagery provides some support for this hypothesis (Tian et al. 2016). Of particular relevance to the question of loudness perception, Tian et al. (2018) found that imagining speaking at different intensities induced neural activity similar to actual perceptual responses and could, moreover, modulate early auditory cortex responses. In order for perceptual estimates to maintain a degree of consistency despite altered feedback, as we see with loudness curves, such internal (or forward) model must nevertheless lie outside of corrective loops, because updated motor plans would alter perception with each correction (cf. Tian and Poeppel 2012).

In sum, the Lombard effect stems from a process that operates independently of conscious control and perception of speech intensity. It is the output of a closed-loop process that adapts the outgoing speech signal based on noise in sensory systems overall. From the standpoint of sensorimotor control, the self-perception of loudness appears as an open-loop process associated with an internal scale that (within the time frames assessed in this study) remains consistent despite significant changes in sensory feedback, as in Lombard speech.

Limitations

A potential confound not controlled for in this study is the possibility of a laryngeal biomechanical response to moderately intense cutaneous vibration of the thyroid lamina. Given the interdependence of biomechanical and neuromuscular systems, a biomechanical or reflexive component to the overall effect is very likely (cf. "tonic vibration reflex"; Bosco et al. 1999). The elicited acoustic patterns do not support a purely biomechanical basis, however. Vibration was associated with a drop in $F2$, suggesting that the pharyngeal cavity of the vocal tract was narrowed and/or that the larynx was raised in this condition. From a biomechanical standpoint, if the presumed cause were a general stiffening of laryngeal musculature, it is unclear why a similar effect did not extend to changes in fundamental frequency (especially given covariation of fundamental frequency and intensity;

Titze 1992). If we consider, furthermore, that the response is sustained over the 10 s of masking, it becomes difficult to attribute the effect entirely to reflex. Ultimately, more explicit testing of biomechanical factors in this experimental paradigm will be needed to properly address this issue.

Another aspect the present study did not control for is subject-specific susceptibility or preference to either auditory or somatosensory masking. A vibration-only condition was not analyzed, because vibration through skin and cartilage affected auditory thresholds when auditory feedback was not masked, confounding auditory and somatosensory effects. Following the methodology adopted by Lametti et al. (2012); however, it could have been possible to evaluate the effect of adding perilaryngeal vibration after auditory masking had been applied. Using this approach, these authors were able to provide a more complete picture of the complementary relationship between auditory and somatosensory feedback in individuals differentially sensitive to one or the other modality. We defer this for future studies employing perilaryngeal vibration.

In conclusion, the masking of laryngeal somatosensory feedback by applying low-pass filtered stochastic vibration to the neck enhanced the Lombard response to high-intensity auditory masking. The qualitative similarities between masking conditions suggest that the Lombard effect is not the result of a specific audio-vocal mechanism, but of a general sensorimotor process that alters speech intensity without affecting the perception or sequencing of planned productions.

Acknowledgements We wish to thank Mark Tiede, Benjamin Elgie and Thomas Gisiger for technical assistance in preparing this experiment.

Compliance with ethical standards

Ethics/declaration Parts of this study were presented in preliminary form at the 2015 Spring meeting of the Acoustical Society of America. The abstract from the proceedings was published in: The Journal of the Acoustical Society of America 137, 2434 (2015); https://doi.org/10.1121/1.4920888.

References

Adelman C, Fraenkel R, Kriksunov L, Sohmer H (2012) Interaction in the cochlea between air conduction and osseous and non-osseous bone conduction stimulation. Eur Arch Otorhinolaryngol 269:425–429

Amazi DK, Garber SR (1982) The Lombard sign as a function of age and task. J Speech Hear Res 25:581–585

Andreatta RD, Barlow SM (2009) Somatosensory gating is dependent on the rate of force recruitment in the human orofacial system. J Speech Lang Hear Res 52:1566–1578

Baken RJ, Orlikoff RF (2000) Clinical measurement of speech and voice, 2nd edn. Cengage Learning, Boston

Bosco C, Cardinale M, Tsarpela O (1999) Influence of vibration on mechanical power and electromyogram activity in human arm flexor muscles. Eur J Appl Physiol 79:306–311

Brajot F, Shiller DM, Gracco VL (2016) Autophonic loudness perception in Parkinson's disease. J Acoust Soc Am 139:1364–1371

Cynx J, Lewis R, Tavel B, Tse H (1998) Amplitude regulation of vocalizations in noise by a songbird (Taeniopygia guttata). Anim Behav 56:107–113

Dean MS, Martin FN (2000) Insert earphone depth and the occlusion effect. Am J Audiol 9:131–134

Fairbanks G (1954) Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. J Speech Hear Disord 19:133–139

Fairbanks G (1960) Voice and Articulation Drillbook, 2nd edn. Harper & Row, New York

Fautrelle L, Bonnetblanc F (2012) On-line coordination in complex goal-directed movements: A matter of interaction between several loops. Brain Res Bull 89:57–64

Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 32:675–701

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

Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 96:280–301

Hammer MJ, Barlow SM (2010) Laryngeal somatosensory deficits in Parkinson's disease: implications for speech respiratory and phonatory control. Exp Brain Res 201:401–409

Hickok G (2012) Computational neuroanatomy of speech production. Nat Rev Neurosci 13:135–145

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Houde JF, Nagarajan SS (2011). Speech production as state feedback control. Frontiers Hum Neurosci. https://doi.org/10.3389/fnhum.2011.00082

Junqua JC (1993) The lombard reflex and its role on human listeners and automatic speech recognizers. J Acoust Soc Am 93(1):510–524

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

Katseff S, Houde J, Johnson K (2011) Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? Lang Speech 55:295–308

Kerrison PD (1918) The tests for malingering in defective hearing. Laryngoscope 28:662–665

Lametti DR, Nasir SM, Ostry DJ (2012) Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. J Neurosci 32:9351–9358

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

Lane HL, Catania AC, Stevens SS (1961) Voice level: autophonic scale, perceived loudness, and effects of sidetone. J Acoust Soc Am 33(2):160–167

Larson CR, Altman KW, Liu H, Hain TC (2008) Interactions between auditory and somatosensory feedback for voice F0 control. Exp Brain Res 187:613–621

Lee BS (1950) Effects of delayed speech feedback. J Acoust Soc Am 22:824–826

Lee GS, Hsiao TY, Yang CCH, Kuo TBJ (2007) Effects of speech noise on vocal fundamental frequency using power spectral analysis. Ear Hear 28(3):343–350

Letowski T, Frank T, Caravella J (1993) Acoustical properties of speech produced in noise presented through supra-aural earphones. Ear Hear 14(5):332–338

Liu H, Wang EQ, Metman LV, Larson CR (2012) Vocal response to perturbations in voice auditory feedback in individuals with Parkinson's disease. PLoS One 7(3):e33629. https://doi.org/10.1371/journal.pone.0033629

Lombard E (1911) Le signe de l'élévation de la voix. Ann des Maladies d'Oreille Larynx Nez Pharynx 37:101–119

Loucks TM, De Nil LF (2012) Oral sensorimotor integration in adults who stutter. Folia Phoniatr Logopaedica 64:116–121

Mu L, Sobotka S, Chen J, Su H, Sanders I et al (2013) Parkinson disease affects peripheral sensory nerves in the pharynx. J Neuropathol Exp Neurol 72(7):614–623

Nonaka S, Takahashi R, Enomoto K, Katada A, Unno T (1997) Lombard reflex during PAG-induced vocalization in decerebrate cats. Neurosci Res 29:283–289

Pick HL Jr, Siegel GM, Fox PW, Garber SR, Kearney JK (1989) Inhibiting the Lombard effect. J Acoust Soc Am 85(2):894–900

Pickett JM (1956) Effects of vocal force on the intelligibility of speech sounds. J Acoust Soc Am 28:902–905

Scheifele PM, Andrew S, Cooper RA, Darre M, Musiek FE et al (2005) Indication of a Lombard vocal response in the St. Lawrence River beluga. J Acoust Soc Am 117(3):1486–1492

Schulman R (1989) Articulatory dynamics of loud and normal speech. J Acoust Soc Am 85(1):295–312

Thierren AS, Lyons J, Balasubramaniam R (2012) Sensory attenuation of self-produced feedback: the Lombard effect revisited. PLoS ONE 7(11):e49370. https://doi.org/10.1371/journal.pone.0049370

Tian X, Poeppel D (2012) Mental imagery of speech and movement implicates the dynamics of internal forward models. Frontiers Psychol. https://doi.org/10.3389/fpsyg.2010.00166

Tian X, Zarate JM, Poeppel D (2016) Mental imagery of speech implicates two mechanisms of perceptual reactivation. Cortex 77:1–12

Tian X, Ding N, Teng X, Bai F, Poeppel D (2018) Imagined speech influences perceived loudness of sound. Nat Hum Behav 2:225–234

Titze IR (1992) Acoustic interpretation of the voice range profile (phonetogram). J Speech Hear Res 35:21–34

Tressler J, Schwartz C, Wellman P, Hughes S, Smotherman M (2011) Regulation of bat echolocation pulse acoustics by striatal dopamine. J Exp Biol 214:3238–3247

Van Summers W, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA (1988) Effects of noise on speech production: Acoustic and perceptual analyses. J Acoust Soc Am 84(3):917–928

Zwislocki JJ, Goodman DA (1980) Absolute scaling of sensory magnitudes: a validation. Percept Psychophys 28:28–38