

Assessing vowel effects on voice quality, and voice quality effects on the respiratory system

2005

Laura L. Koenig^{1,a)} and Susanne Fuchs²¹Department of Communication Sciences and Disorders, Adelphi University, Garden City, New York 11530, USA²Leibniz-Centre General Linguistics (ZAS), Berlin 10117, Germany

koenig@haskins.yale.edu, fuchs@leibniz-zas.de

Abstract: This study assesses (a) effects of vowel height and tense-lax status on the laryngeal closed quotient (CQ) and (b) whether respiratory volume changes vary with differences in CQ. German speakers produced words containing eight different vowels in normal and loud conditions. The only significant vowel effect was found for the /a:-a/ pair, with lower CQ in /a/ at normal intensity. There was an insignificant trend for lower CQ to be associated with more negative thoracic slopes. The CQ difference for the /a:-a/ contrast, which relies more on duration than other tense-lax pairs in German, requires further study. © 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Brad H. Story]

<https://doi.org/10.1121/10.0003533>**Received:** 24 September 2020 **Accepted:** 25 January 2021 **Published Online:** 24 February 2021

1. Introduction

Research has documented numerous interactions among respiratory, laryngeal, and supralaryngeal components of the speech system. In this paper, we evaluate how vowel quality, specifically height and tense-lax differences, may affect voice quality, as measured by electroglottography (EGG). We also assess relationships between voice quality and respiratory system measures and consider whether they change in louder speech.

1.1 Supralaryngeal–laryngeal interactions

Supraglottal configurations may affect vocal-fold vibration in various ways, e.g., via acoustic coupling or mechanical forces acting on laryngeal structures. The “intrinsic f₀” effect, whereby higher vowels tend to have slightly higher values of fundamental frequency (f₀), is a well-established example of how vowel postures affect phonation [e.g., Whalen and Levitt (1995)]. Vowel effects on other laryngeal measures are less clear.

Of particular interest in this paper is whether vowel qualities correlate with a measure of voice quality that relates to the degree of breathiness in a signal. Voice quality can be measured in various ways. In airflow and/or EGG signals, a common measure is the open quotient or its inverse, the closed quotient (CQ). Some authors have also used measures of spectral tilt, comparing the relative amount of energy in the first harmonic (H1) to some higher-frequency reference. Higher open quotients, lower CQs, and steeper spectral tilts correspond perceptually to breathier voice qualities (Childers and Lee, 1991).

Experimental and modeling data from Bickley and Stevens (1986) suggested that more extreme supraglottal constrictions lead to slower glottal closing and more sinusoidal pulse shapes, other characteristics of breathy vocal quality. Some studies have reported data that support this hypothesis for consonants [see Chong *et al.* (2020) and references therein]. The prediction can be extended to considerations of vowel height; it could also hold for tense-lax pairs insofar as that contrast is reflected in vowel height differences.

Empirical evidence for correspondences between vowel quality and voice quality are sparse and contradictory, however. Esposito *et al.* (2019) pointed out that measures of linguistically contrastive phonation types have usually been restricted to low vowels.¹ In their cross-linguistic study covering many vowel types, these authors found that breathier phonation, as measured by the amplitudes of the first and second harmonics, correlated with lower values of both the first and second formants (F1, F2). The F1 results are consistent with the Bickley and Stevens model, but those for F2 are unexpected. In a study of possible secondary cues for an apparent vowel merger in a dialect of American English, Di Paolo and Faber (1990) observed a voice quality difference between high and mid tense-lax pairs. The authors reported breathier voice qualities in the tense members of the pairs as measured by the relative amplitudes of H1 and F1, i.e., a voice quality difference in the predicted direction. On the other hand, in an EGG study on German, Marasek (1997)

^{a)} Author to whom correspondence should be addressed. Also at: Haskins Laboratories, New Haven, CT 06511, USA.

reported the opposite pattern for tense-lax pairs: In stressed vowels, open quotient values were slightly higher in lax vowels compared to tense ones. These results were complicated, however, by interactions with stress and word repetition, as well as differences between men and women and among the five tense-lax pairs. In short, the limited empirical data do not provide a clear picture of whether there may be systematic voice quality variation as a function of vowel quality. It is possible that varying results reflect language-specific differences, i.e., some correlations between vowel quality and voice quality could be learned behaviors instead of an automatic consequence of laryngeal–oral coupling.

1.2 Laryngeal–respiratory interactions

When the glottis is open for voiceless consonants, rapid venting of air may be reflected in respiratory system data [see [Fuchs et al. \(2019\)](#) and references therein]. It is conceivable that more modest differences in laryngeal configuration may also affect respiratory measures. Indeed, such suggestions have been made in the clinical voice literature. For example, [Sapienza and Stathopoulos \(1994\)](#) observed both higher oral airflow rates, indicating breathier voice qualities, and larger lung volume excursions in women and children with bilateral vocal-fold nodules as compared to control speakers without nodules.

1.3 Loudness manipulations

Loudness, or increased vocal effort, represents another case of laryngeal–respiratory interaction: Louder speech, which is primarily ascribed to changes in respiratory system pressures [e.g., [Ladefoged and McKinney \(1963\)](#)] is associated with greater vocal-fold contact ([Orlikoff, 1991](#)), via changes in aerodynamic forces, laryngeal muscle activity ([Finnegan et al., 2000](#)), or some combination of the two. An aerodynamic modeling study by [Zhang \(2016\)](#) also points to a possible role for laryngeal adjustments in regulating respiratory effort. Louder or more effortful speech can have some effects on supralaryngeal articulation as well [see review in [Koenig and Fuchs \(2019\)](#)]. It is therefore of interest to ask whether any laryngeal–supralaryngeal interactions observed in normal speech carry over to loud speech.

1.4 Summary and hypotheses

To explore how vowel qualities may relate to voice qualities, and whether any voice quality variation may relate to respiratory behavior, we drew on data from German, which has a large vowel inventory that includes multiple heights and tense-lax variation. The tense-lax contrasts are mostly manifested as quality differences, i.e., differences in F1 and F2 values. However, the low tense-lax pair can be an exception: among others, [Pompino-Marschall \(2009\)](#) reports that in Northern Standard German /a/ and /a:/ are distinguished by quantity (long vs short) rather than quality (spectral differences).

The first goal of our study was to evaluate whether German vowel qualities are consistently associated with a measure of voice quality, namely the CQ in the EGG signal. Second, we asked whether voice quality variation is reflected in respiratory system measures. Following [Bickley and Stevens \(1986\)](#) and the acoustic studies reported in Sec. 1.1, we hypothesized that higher vowels have lower CQs. This pattern should also hold for tense-lax vowel pairs in which the tense member is higher in vowel space. We further hypothesized that a lower CQ correlates with a steeper declining slope of the thorax, i.e., thoracic volume decreases more quickly when glottal closure is less complete. This follows the results of [Fuchs et al. \(2019\)](#), where thoracic slopes differed between open vs closed glottal conditions (/p/ vs /b, m/), but no differences were observed for abdominal measures (presumably because the abdomen is a greater distance from the glottis).

2 Methodology

2.1 Speakers and recording

All 13 participants were female native speakers of German, ages 19–37 yrs (mean = 24.7 yrs), body mass indices 17.8–24.2 (mean = 20.04). To limit upper body movement, participants were recorded in a seated position. Three signals were recorded simultaneously: thoracic displacements (obtained using inductive plethysmography); acoustics (Sennheiser microphone HKH50 P48, Wedemark, Germany); and EGG (Glottal Enterprises EG2-PCX, Syracuse, New York). Speakers produced bisyllabic target words in a question-answer task that gave participants a fixed prosodic and syntactic framework. For example:

Q: Magst du München? [*Do you like Munich?*]

A: München mag ich, aber nicht Hamburg [*I like Munich, but not Hamburg*].

This paradigm ensured that all target words were produced utterance-initially and under sentential focus. Participants supplied their own word at the end of the sentence (above, *Hamburg*) to encourage active participation in the task.

Words had initial labial consonants /m b p/ and a medial alveolar obstruent, and one of the vowels /i ɪ y ʏ u a: a:/ for /i/, “Miete” (*rent*) and “Pita” (*pita*); for /ɪ/, “Mitte” (*section of Berlin*) and “Pizza” (*pizza*); for /y/, “Büsten” (*busts*) and “Büsum” (*a city*); for /ʏ/, “Mützen” (*caps*) and “München” (*Munich*); for /u/, “Pudel” (*poodle*) and “Pute” (*turkey*); for /ʊ/, “Butter” (*butter*) and “Pudding” (*pudding*); for /a:/ (tense), “Mate” (*a tea*) and “Paten” (*godparents*); for /a/ (lax), “Paddeln” (*to canoe*) and “Pasta” (*pasta*). There were thus six high vowels and two low vowels. Words were elicited five times in randomized order. Utterances were produced in normal and loud conditions. For the loud condition, the experimenter moved to another

room, behind a window, and spoke loudly through the glass. Participants chose their own degree of loudness. The comparison between loud and normal speech allowed us to assess whether any vowel differences would hold across this manipulation, which, as noted above, can affect laryngeal as well as supralaryngeal measures.

2.2 Processing and measures

Vowels were marked in PRAAT (Boersma and Weenink, 2017) from the release of the initial consonant (for voiced sounds) or the onset of voicing (for aspirated sounds) to the closure of the medial consonant. Automatic scripts were used to obtain first formant values (F1) at the temporal midpoint of all vowels with subsequent correction in case of clearly mis-tracked formants. The F1 covaries with height differences as well as the tense-lax distinction: High lax vowels such as /ɪ/ are generally expected to have higher F1s than their tense counterparts (e.g., /i/). For the low tense-lax pair, the F1s could be equivalent, following Pompino-Marschall (2009), or else the lax vowel /a/ might show lower values of F1 than the tense /a:/, i.e., be more of a mid vowel than a low one.

The same acoustic onset and offset values were used to extract a window of 40 ms centered at the midpoint of each vowel in both physiological signals (thorax and EGG). The 40 ms window was chosen to limit consonantal influences somewhat, while still retaining most of the data (ca. 93.6% of vowel productions).

The EGG excerpts were detrended, smoothed, and first-differenced in MATLAB. Events were then labeled automatically as follows. (a) The peak in the difference signal was taken as the time of glottal closure. (b) Periods were defined based on the adjacent first difference peaks. (c) The time of glottal opening was defined as 3/7 of the maximum in the EGG signal to the following minimum [see Henrich *et al.* (2004)]. Given that the EGG signal provides a clearer indication of the vocal-fold contact than the degree of opening, we used here the CQ measure. The CQ was defined as the duration of the closed phase divided by the period duration.

Also, in MATLAB, the respiratory signals were measured automatically by taking the slope of the thorax displacement over the 40 ms excerpt. Slope was defined in the typical way as a difference in y divided by a difference in x , i.e., the difference in magnitude of thoracic values divided by 40 ms. The full dataset for the two analyses, across speakers and conditions, consisted of $N = 1854$ tokens.

For all statistical analysis, we used R (R Core Team, 2020). The GGLOT2 library was used for data visualization, the LME4 package for running linear mixed effect models, and EMMEANS for *post hoc* comparisons with the Tukey method for adjustment of p -values. In a first linear mixed effect model, the CQ served as the dependent variable; tenseness (lax vs tense vowels), height (high vs low vowels), and condition (normal vs loud) and their interactions were included as fixed effects. We chose the tense, high vowel in normal loudness as the reference level. The second linear mixed effect model used thoracic slope as the dependent measure. Since our main prediction here was that more open glottal configurations would be associated with steeper slopes, the vowels were pooled, and the CQ and loudness served as the fixed effects. We began by including speaker-specific random slopes for all fixed effects and a random effect for each file in both analyses. The random factor of file was to control for potential changes in behavior over the course of the experiment. The model comparison showed that the model without the file factor had the lower AIC value (Akaike information criterion); thus, that factor was removed for the final analyses.

3. Results

3.1 Vowel quality and loudness effects on the CQ

The linear mixed effect models revealed no significant main effects, but only a trend for a slightly higher CQ in loud speech ($p = 0.0696$). There were multiple interactions. Only one comparison turned out to be significant in the *post hoc* analyses: tense and lax low vowels differed in normal speech [$\beta = 0.0277$, standard error (SE) = 0.00722, $t = 3.834$, $p = 0.0075$]. As shown in Fig. 1, the lax low vowel was produced with a lower CQ (CI: 0.314–0.359) than its tense counterpart (CI: 0.342–0.387). It should be noted that, whereas both words for /a/ began with a voiceless aspirated stop (*Pasta, paddeln*), the words for /a:/ included both an aspirated stop and a nasal (*Paten, Mate*). Removing the word *Mate* from the analysis weakened the effect but did not remove it ($\beta = 0.02632$, SE = 0.00823, $t = 3.196$, $p = 0.0379$).² The difference, therefore, does not seem to reflect context effects.

3.2 Voice quality and loudness effects on respiratory data

The model predicting the thorax slope based on the CQ and loudness revealed no main effects and no interactions. There was an insignificant trend ($p = 0.0653$) that more negative (steeper) thorax slopes coincided with lower CQ values, leading to a positive relationship. There was also a trend ($p = 0.0916$) for steeper slopes in louder speech. Figure 2 shows that the (weak) pattern for CQ is evident in normal speech, but not loud. The data show considerable variability, particularly in loud speech.

4. Discussion and conclusions

The lack of a main effect of height and the absence of tense-lax differences in the high vowel pairs speak against vowel quality yielding automatic voice quality differences in speakers of German, as would be predicted by the Bickley and Stevens (1986) model. The only finding was a lower CQ for the lax vowel /a/, which is higher in vowel space than its tense

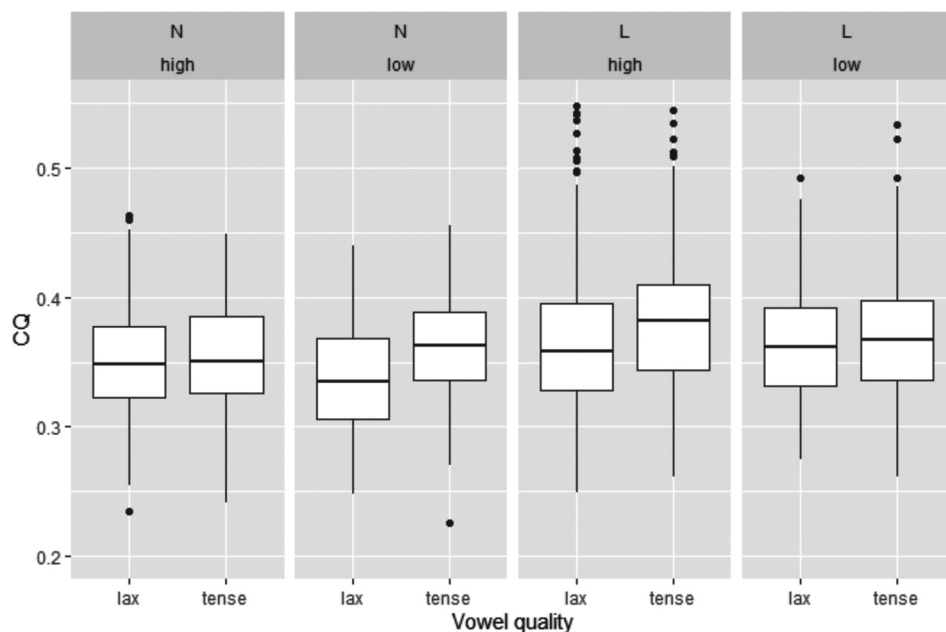


Fig. 1. Boxplots [median values as horizontal bold lines, whiskers as 95% confidence interval (CI), and dots for outliers] showing CQ for the three different factors: loudness [normal (N) vs loud (L)]; vowel height (high vs low); and tenseness (lax vs tense).

counterpart /a:/.³ Our results are somewhat consistent with Marasek (1997), but more specific. That study reported some overall effects of tense vs lax quality, with higher open quotients in lax vowels, whereas for our data, the voice quality difference was limited to /a-a:/.

As noted in Sec. 1.1, the distinction between /a/ and /a:/ in Northern Standard German has been claimed to rely primarily on length (Pompino-Marschall, 2009), with little spectral difference. To assess this possibility, we performed *post hoc* analyses on the midpoint F1 values and the vowel durations. Linear mixed models were calculated as for CQ, but with F1 or duration as dependent variables. Results for F1 revealed significant differences for all four tense-lax vowel pairs (all $p < 0.001$). For normal speech, pairs differed by about 100 Hz: For high vowels, $\beta = -104$ Hz, $t = -11.95$, and for low vowels, $\beta = 113$ Hz, $t = 11.4$. The sign of the effect differs, i.e., for the low vowels, the lax member of the pair is higher in vowel space. For loud speech, the F1 difference is smaller for low vowels than high vowels (high vowels: $\beta = -194$ Hz, $t = -22.86$; low vowels: $\beta = 72$ Hz, $t = 7.12$). For duration, the tense-lax differences were all significant (all $p \leq 0.001$), but larger for the low vowels: In normal speech, $\beta = 25$ ms, $t = 6.16$ for high vowels, and $\beta = 61$ ms, $t = 13.96$ for low vowels; in loud speech, $\beta = 34$ ms, $t = 8.55$ for high vowels vs $\beta = 63$ ms, $t = 14.54$ for low vowels. As expected, the /a/-/a:/ distinction relies heavily on duration, but there is some spectral difference as well. Similar results have recently been reported by Gao *et al.* (2020).

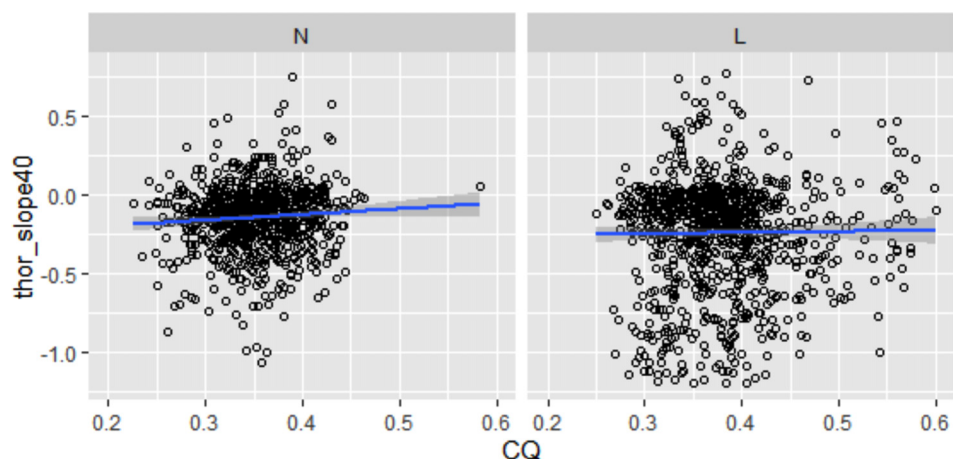


Fig. 2. Scatterplots for CQ (x) and thoracic slope (y). Linear regression lines are included with gray areas for confidence bands. Data are split by condition (normal vs loud).

It is possible that the low tense-lax pair, different from other tense-lax pairs in German by virtue of its reliance on duration, is partly differentiated by vowel quality. This would be conceptually analogous to the findings of Di Paolo and Faber (1990), but their data showed breathier voice quality in the tense members of the pairs (considering high and mid vowels only). The direction of our effect is consistent with that of Marasek (1997) for German, but not as general. The nature of the /a/-/a:/ contrast in German appears to call for further study, and it may be relevant to assess dialectal variation as well. To determine whether German listeners make use of any voice quality differences between /a/ and /a:/ as a secondary cue for the phonological contrast, would, of course, require perceptual testing.

The second hypothesis, that differences in CQ would relate to respiratory system data, was not confirmed. There was only a nonsignificant trend for steeper (more negative) thoracic slopes with lower values of CQ, which, qualitatively, was evident in normal speech but not loud. As indicated earlier, some studies assessing laryngeal pathologies have observed effects of degree of glottal closure on breathing data [e.g., Sapienza and Stathopoulos (1994)]. Possibly, the differences in laryngeal configuration obtained for the normal speakers in our study were not of sufficient magnitude to yield measurable effects using inductance plethysmography.

In sum, the data do not confirm either of the hypotheses regarding automatic interactions among laryngeal and supralaryngeal systems or laryngeal and respiratory systems. Respiratory measures did not show a clear relationship with voice quality, and voice quality only differed for the low vowel pair (/a/ vs /a:/). The lack of a general height effect runs counter to the predictions made by the Bickley and Stevens (1986) model. These data, combined with the conflicting results obtained by past work, suggest that effects of vowel quality on voice quality may be language-specific rather than automatic. The results obtained for the low vowel pair should be confirmed in subsequent studies of German speakers across various dialectal backgrounds. Additionally, the possible perceptual utility of these voice quality differences should be verified in listening studies that control for durational and spectral differences between the vowels.

Acknowledgments

This work was supported by a grant from the Leibniz Association to ZAS. We thank Jörg Dreyer for technical support, Moriah Rastegar and Thomas Moran for assistance with manuscript preparation (supported by Research Assistantships from Adelphi University), and our speakers for their participation. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

References and links

¹This situation reflects methodological factors. In linguistic fieldwork, acoustic recordings are far more common than physiological measures. When assessing voice quality using acoustic measures of spectral tilt, a correction factor needs to be applied to higher vowels to adjust for effects of F1 on lower harmonic amplitudes. Such concerns do not apply to physiological data.

²As a final check on the possibility that coarticulatory influences could account for the results, we graphed the CQ values over the full 40 ms time windows for both /a/ and /a:/. Only one of the 13 speakers showed a pattern whereby the CQ changed early in the window for /a/, converging toward that of /a:/, as would be expected if carryover effects of aspiration were involved. Again, therefore, it does not seem that the /a/-/a:/ difference in our data can be attributed to phonetic context effects.

³An examination of Fig. 1 suggests that loudness effects may play out differently for high and low vowels. In louder speech, CQ for lax /a/ became more similar to tense /a:/, i.e., the vowel difference disappeared. For the high vowels, tense and lax vowels had comparable CQ values in normal speech, whereas in loud speech, they show a qualitative difference (not significant in our analyses). There is reason to think that vowel qualities and loudness interact. Koenig and Fuchs (2019) assessed vowel formants in normal and loud speech in German women, many of whom were included in the current dataset. Whereas lax and low vowels were lower in louder speech (higher F1 values), the high tense vowels did not show F1 changes. Thus, if there is a tendency for high tense vowels to have higher CQ in loud speech, it arises in the presence of stable F1s, counter to predictions of the Bickley and Stevens model.

Bickley, C. A., and Stevens, K. N. (1986). "Effects of a vocal-tract constriction on the glottal source: Experimental and modelling studies," *J. Phon.* **14**, 373–382.

Boersma, P., and Weenink, D. (2017). "Praat: Doing phonetics by computer (version 6.0.36) [computer program]," <https://www.fon.hum.uva.nl/praat/> (Last viewed 8 February 2021).

Childers, D. G., and Lee, C. K. (1991). "Vocal quality factors: Analysis, synthesis, and perception," *J. Acoust. Soc. Am.* **90**, 2394–2410.

Chong, A. J., Risdal, M., Aly, A., Zymet, J., and Keating, P. (2020). "Effects of consonant constrictions on voice quality," *J. Acoust. Soc. Am.* **148**, EL65–EL71.

Di Paolo, M., and Faber, A. (1990). "Phonation differences and the phonetic content of the tense-lax contrast in Utah English," *Lang. Var. Change* **2**, 155–204.

Esposito, C. M., Sleeper, M., and Schäfer, K. (2019). "Examining the relationship between vowel quality and voice quality," *J. Int. Phon. Assoc.* 1–32.

Finnegan, E., Luschei, E. S., and Hoffman, H. T. (2000). "Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech," *J. Speech Lang. Hear. Res.* **43**, 934–950.

Fuchs, S., Koenig, L. L., and Petrone, C. (2019). "Exploring the source of short-term variations in respiratory data," *J. Acoust. Soc. Am.* **145**, EL66–EL71.

Gao, Y., Ding, H., and Birkholz, P. (2020). "An acoustic comparison of German tense and lax vowels produced by German native speakers and Mandarin Chinese learners," *J. Acoust. Soc. Am.* **148**, EL112–EL118.

- Henrich, N., d'Alessandro, C., Castellengo, M., and Doval, B. (2004). "On the use of the derivative of electroglottographic signals for characterization of nonpathological voice phonation," *J. Acoust. Soc. Am.* **115**, 1321–1332.
- Koenig, L. L., and Fuchs, S. (2019). "Vowel formants in normal and loud speech," *J. Speech Lang. Hear. Res.* **62**, 1278–1295.
- Ladefoged, P., and McKinney, N. (1963). "Loudness, sound pressure, and subglottal pressure in speech," *J. Acoust. Soc. Am.* **35**, 454–460.
- Marasek, K. (1997). "Electroglottographic description of voice quality," Habilitation, University of Stuttgart, Stuttgart, Germany.
- Orlikoff, R. F. (1991). "Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects," *J. Speech Hear. Res.* **34**, 1066–1072.
- Pompino-Marschall, B. (2009). *Einführung in die Phonetik (Introduction to Phonetics)* (Walter de Gruyter, Berlin).
- R Core Team (2020). "R: A Language and Environment for Statistical Computing (version 4.0.3)," R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (Last viewed 10 October 2020).
- Sapienza, C. M., and Stathopoulos, E. T. (1994). "Respiratory and laryngeal measures of children and women with bilateral vocal fold nodules," *J. Speech Lang. Hear. Res.* **37**, 1229–1243.
- Whalen, D. H., and Levitt, A. G. (1995). "The universality of intrinsic f₀ of vowels," *J. Phon.* **23**, 349–366.
- Zhang, Z. (2016). "Respiratory laryngeal coordination in airflow conservation and reduction of respiratory effort of phonation," *J. Voice* **30**(6), 760.e7–760.e13.