Elicitation context does not drive F0 lowering following voiced stops

Citation for published version:
Kirby, J, Ladd, B, Gao, J & Elliott Slosarova, Z 2020, 'Elicitation context does not drive F0 lowering following voiced stops: Evidence from French and Italian', *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. EL147-EL152. https://doi.org/10.1121/10.0001698

Digital Object Identifier (DOI):
10.1121/10.0001698

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
The Journal of the Acoustical Society of America

Elicitation context does not drive F0 lowering following voiced stops: Evidence from French and Italian

James Kirby, D. Robert Ladd, Jiayin Gao, and Zuzana Elliott

Citation: The Journal of the Acoustical Society of America **148**, EL147 (2020); doi: 10.1121/10.0001698
View online: https://doi.org/10.1121/10.0001698
View Table of Contents: https://asa.scitation.org/toc/jas/148/2
Published by the Acoustical Society of America

JASA EXPRESS LETTERS

https://doi.org/10.1121/10.0001698

Elicitation context does not drive F0 lowering following voiced stops: Evidence from French and Italian

James Kirby,[a] D. Robert Ladd, Jiayin Gao, and Zuzana Elliott
Department of Linguistics and English Language, University of Edinburgh, Edinburgh, EH8 9AD, United Kingdom
j.kirby@ed.ac.uk, d.r.ladd@ed.ac.uk, jiayin.gao@ed.ac.uk, zuzana.elliott@ed.ac.uk

Abstract: Consonant-intrinsic F0 (CF0) effects are mainly the result of raising F0 following voiceless obstruents, rather than of lowering F0 following voiced obstruents. However, there are also documented instances where lowered F0 following voiced obstruents is enhanced. Given that both voicing and F0 are affected by prosodic context, it is possible that CF0 is lowered in some contexts but not others. This possibility is investigated by examining CF0 in French and Italian in isolated citation forms. Results are comparable to carrier-phrase contexts, where no F0 lowering after voiced obstruents is observed. Possible sources of the apparent cross-linguistic differences are discussed.

[Editor: Rachel M. Theodore] Pages: EL147–EL152

Received: 8 April 2020 **Accepted:** 17 July 2020 **Published Online:** 18 August 2020

1. Introduction

In many (possibly all) languages, obstruent consonants influence the fundamental frequency (F0) of the following vowel. We refer to this microprosodic effect as *consonant-intrinsic F0* (CF0). Hanson (2009) found that post-pausal voiceless obstruents in American English have F0-raising properties, while CF0 is conditioned by phonologically voiced obstruent patterns with sonorants. Subsequent studies of languages including French and Italian (Kirby and Ladd, 2016), Thai, Khmer, and Nothern Vietnamese (Kirby, 2018), and Tokyo Japanese (Gao and Arai, 2019) all report little or no difference in the F0 of vowels following sonorants and those following phonologically (and phonetically) voiced plosives. Together with the work of Löfqvist *et al.* (1989), such findings led Hanson (2009) and Kirby and Ladd (2016) to argue that CF0 effects are driven primarily by a gesture that induces F0 *raising* following voiceless segments, rather than F0 *lowering* following voiced segments (cf. Mohr, 1971; Hombert, 1978; Kingston and Diehl, 1994).

However, researchers have also documented cases where F0 following voiced obstruents is lower than that of sonorants. In Afrikaans, Coetzee *et al.* (2018) show that CF0 is higher after nasals than after phonologically voiced plosives for the majority of the vocalic excursion, regardless of whether or not the plosive was realized with or without closure voicing. Similar findings are reported by Howe (2017) for Central Malagasy. In both languages, F0 appears to have become the primary cue to the laryngeal contrast, replacing a historical lag/lead dichotomy nominally similar to that found in languages such as French and Italian.

Precisely how such shifts take place remains debated. One possibility is that the perceptual salience of the CF0 dichotomy increases when speakers fail to overcome the aerodynamic constraints necessary for voicing and instead produce devoiced stops. When exposed to such talkers, some listeners might then increase the perceptual cue weight of the F0 differences. If the CF0 dichotomy is primarily the result of a laryngeal maneuver accompanying the production of voiceless plosives, and if it is this gesture that speakers target to enhance the CF0 dichotomy in production, there is no reason to expect sonorants to be affected, and hence, F0 following sonorants and devoiced plosives would presumably remain similar. The fact that F0 is demonstrably lower when following voiced plosives than sonorants in Afrikaans and Central Malagasy suggests that, at least in some cases, enhancement of CF0 can focus on the perturbations following [+voice] as well as [–voice] segments (Kingston and Diehl, 1994). This leads us to ask if there exist conditions under which F0 following [+voice] obstruents is naturally lowered, relative to sonorants, which might increase its perceptual salience to listeners as a target for further enhancement.[1]

One possible conditioning environment is differences in prosodic context. In the studies of Hanson (2009) and Kirby and Ladd (2016), where post-release CF0 for [+voice] stops was

[a]Author to whom correspondence should be addressed, ORCID: 0000-0002-0502-5245.

JASA EXPRESS LETTERS

found to pattern with sonorants, utterances were embedded in a carrier phrase with varying positional and prosodic contexts (strong vs weak focus, post-pausal vs intersonorant, phrase-initial vs phrase-medial, etc.). In Kirby and Ladd (2016), for example, the words containing the target syllables were at the right edge of a prosodic domain, but the position of the target syllable within the word was not carefully controlled. In the study of Coetzee *et al.* (2018), the target sequence was always word-initial and post-pausal; items were presented in list format and were therefore read with a typical list intonation (Coetzee, 2019). In typical list intonation, a low accentual target is arguably followed by a high boundary tone, and if the consonantal gesture and pitch target reinforce one another, F0 could be lowered as a result. Minimally, this means that direct comparison between the results of Coetzee *et al.* (2018) and studies in which no difference between CF0 following sonorants and [+voice] plosives was found has to be done with caution, but it suggests the possibility that elicitation context may have an effect on CF0 following [+voice] segments.

To determine whether these methodological and prosodic differences can in fact impact the realization of CF0, we conducted a study of French and Italian citation forms to allow us to compare more directly with the results of Coetzee *et al.* (2018), as well as to compare with the carrier phrase contexts of Kirby and Ladd (2016).

2. Methods

2.1 Materials

Coetzee *et al.* (2018) analyzed 60 items, grouped into minimal or near-minimal mono- or disyllabic pairs beginning with /p b/ or /t d/ (15 with each onset). As a comparable list would have been impossible to construct for Italian (as monosyllabic words are largely restricted to function words and recent loans), we instead sought out cognate or near-cognate items in French and Italian to minimize the nuisance effects of token differences. In Italian, items differed in whether the initial syllable bore lexical stress, while in French, we contrasted monosyllables with disyllables. We refer to these as the TONIC and PRETONIC conditions, respectively. The target sequence (/ba/, /pa/, or /ma/) was always word-initial, as in Coetzee *et al.* (2018). The word list included 12 words with each target sequence in each position, for a total of 72 items in each language (see supplementary materials,[2] Appendix A). An additional 72 items (containing vowels other than /a/ and coronal or velar onsets) were included as distractors.[3]

2.2 Participants

Eleven native speakers of French (9 female) and 10 native speakers of Italian (7 female) were recruited from the local community in Edinburgh, UK. These participants were staying in Edinburgh temporarily, either as visitors or as exchange students, and none identified as immigrants. We restricted participation to those who had completed both primary and secondary school in France or Italy, as appropriate. Mean age was 21 (SD = 1.5) for French participants and 31 (SD = 10.8) for Italians. We preferred participants who had spent less time in the UK, but this was not always possible to control for. For the French participants, mean time in the UK at time of recording was 5.6 months, while for Italians it was 19 months.

To facilitate comparison to Coetzee *et al.* (2018), and because we do not have any hypotheses about the possible effects of speaker sex, here we focus on the 16 female participants. However, the data from the male speakers is included in the supplementary materials.

2.3 Recording and data processing

Participants were recorded individually in a soundproof testing booth at the University of Edinburgh using a Shure SM7 cardioid pattern microphone and the SpeechRecorder software (Draxler and Jänsch, 2004). Each participant received the same set of instructions, which outlined the task procedure but did not divulge the purpose of data collection or identify the target tokens we were analyzing. Although participants' interactions with the experimenter were carried out in English, all instructions, stimuli, and break screens during the recording were written in French or Italian, as appropriate.

After a brief warm-up, participants were recorded reading each item. The target items and distractors were randomized and organized into five blocks, and participants were encouraged to take a short break after each block. In total, each recording session took no longer than 20 min.

Target utterances were then manually segmented to indicate the extent of voice onset time (VOT) and the duration of the post-release vowel, and stored as an EMU speech database (Winkelmann *et al.*, 2017). F0 estimates were made at 5 ms intervals using the ksvf0 pitch estimator of the wrassp package (Bombien *et al.*, 2018). F0 was calculated across all voiced frames of each CV sequence (including the closure phase for /ba/ and /ma/ sequences).

JASA EXPRESS LETTERS
https://doi.org/10.1121/10.0001698

3. Results

3.1 F0 profiles

Figure 1 shows the mean speaker-scaled F0 contours over the CV sequences of interest by language and onset type in both syllable positions. To test the hypothesis that F0 following /b/ would be lower than that of /m/, we fit two Bayesian hierarchical linear models to Hertz values for each language separately, using the brms package (Bürkner, 2017). One model was fit to F0 averaged over $\pm 10\%$ of the closure midpoints, and the other to F0 values averaged over the first 25% of the vowel. All models included the same covariate structure, with treatment-coded predictors ONSET, SYLLABLE, and their interaction. All models also included random intercepts for items and speakers, and by-speaker random slopes for ONSET, SYLLABLE, and their interaction. For each model, the baseline (intercept) represents French speakers' mean F0 (in Hz) for /m/ in pretonic position.

For most parameters, we set weakly regularizing normal or half-normal priors centered at zero, but the prior for ONSET was centered around 20 with a standard deviation of 50. The standard deviation estimate comes from our own previous work on onset F0 in a range of languages (Kirby and Ladd, 2016; Kirby, 2018; Gao and Arai, 2019). The mean of 20 was selected based on Coetzee *et al.* (2018), as this appears to be roughly the mean difference (in Hertz) between nasals and plosives (voiceless plosives being on average about 20 Hz higher, and voiced plosives about 20 Hz lower). We also fit models with a weakly informative prior for ONSET, but in the end, the model estimates proved very robust to choices of priors.

Four sampling chains ran for 2000 iterations with a warm-up period of 1000 iterations for each model, thereby yielding 4000 samples for each parameter tuple. For the relevant cell means and differences between them, we report the expected values under the posterior distribution and their 95% credible intervals (CIs). For differences between cells, we also report the posterior probability that a difference δ is greater than zero. Full model output and plots for individual speakers are available in the supplementary materials (Appendix B).

Much as in Kirby and Ladd (2016), lowering of F0 is observed during the closure phase for /b/ compared to /m/. There is weak evidence that F0 lowering during the closure is greater in Italian than in French $(\mathbb{E}(\mu_{\text{French}\,/b/}) - \mathbb{E}(\mu_{\text{Italian}\,/b/}) = 10, \text{CI} = [-17, 36], P(\delta > 0) = 0.78)$, but this is partly due to the fact that the slope of F0 for both /m/ and /b/ is negative during the closure for French, whereas the F0 contour of /m/ during the closure is flatter for Italian. Averaging over syllable position, F0 for /b/ is about 8 Hz lower at midpoint than for /m/ in French $(\mathbb{E}(\mu_{/m/}) - \mathbb{E}(\mu_{/b/}) = 8, \text{CI} = [0, 15])$, roughly half as much as in Italian $(\mathbb{E}(\mu_{/m/}) - \mathbb{E}(\mu_{/b/}) = 15, \text{CI} = [2, 28]$; both $P(\delta > 0) = 0.98)$.

In the first 25% of the post-release phase, the average F0 in pretonic position is higher compared to tonic position in both languages (French: $\mathbb{E}(\mu_{\text{pretonic}}) - \mathbb{E}(\mu_{\text{tonic}}) = 6, \text{CI} = [1, 10], P(\delta > 0) = 0.99$; Italian: $\mathbb{E}(\mu_{\text{pretonic}}) - \mathbb{E}(\mu_{\text{tonic}}) = 10, \text{CI} = [-3, 21], P(\delta > 0) = 0.96)$.

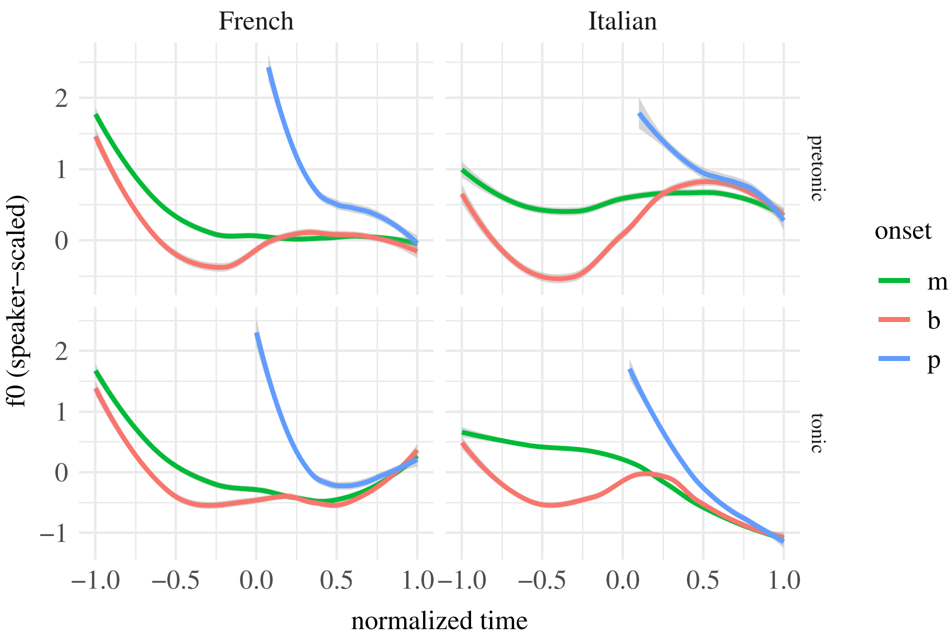

Fig. 1. (Color online) Average time- and speaker-normalized pitch contours (loess smoothing) during period of oral closure (–1 to 0) and following release (0 to 1).

JASA EXPRESS LETTERS

https://doi.org/10.1121/10.0001698

Table 1. Posterior estimates and 95% confidence intervals of effect on F0 (Hz) averaged over first 25% of vowel. Intercept for each language represents the estimate for /m/ in pretonic position.

	Estimate	Est. Error	Q2.5	Q97.5
Language = French				
Intercept	192.04	17.32	156.55	226.65
Onsetb	1.31	1.51	–1.71	4.20
Onsetp	19.21	2.43	14.31	24.01
Syllabletonic	–4.20	2.02	–8.26	–0.33
Onsetb:syllabletonic	0.00	2.02	–3.88	3.94
Onsetp:syllabletonic	–2.57	2.09	–6.68	1.62
Language = Italian				
Intercept	183.25	21.46	139.59	225.17
Onsetb	–0.75	1.73	–4.13	2.65
Onsetp	14.44	2.36	9.61	19.15
Syllabletonic	–7.10	3.78	–14.40	0.46
Onsetb:syllabletonic	1.72	2.13	–2.41	5.85
Onsetp:syllabletonic	–1.39	2.33	–5.95	3.16

The critical within-language pairwise comparisons are given in Table 1, showing that the differences in the expected values under the posterior distribution for post-release F0 between /m/ and /b/ are negligible in both languages.

3.2 Voice onset time

For completeness, we include a brief overview of the VOT distributions. Figure 2 shows histograms of VOT by language and syllable position, with summary statistics given in Table 2. Full model output and plots for individual speakers are available in the supplementary materials (Appendix C).

Mean differences in VOT duration between languages were negligible (5–8 msec). VOT for /b/ in tonic syllables is around 16 msec longer than in pretonic syllables in both French and

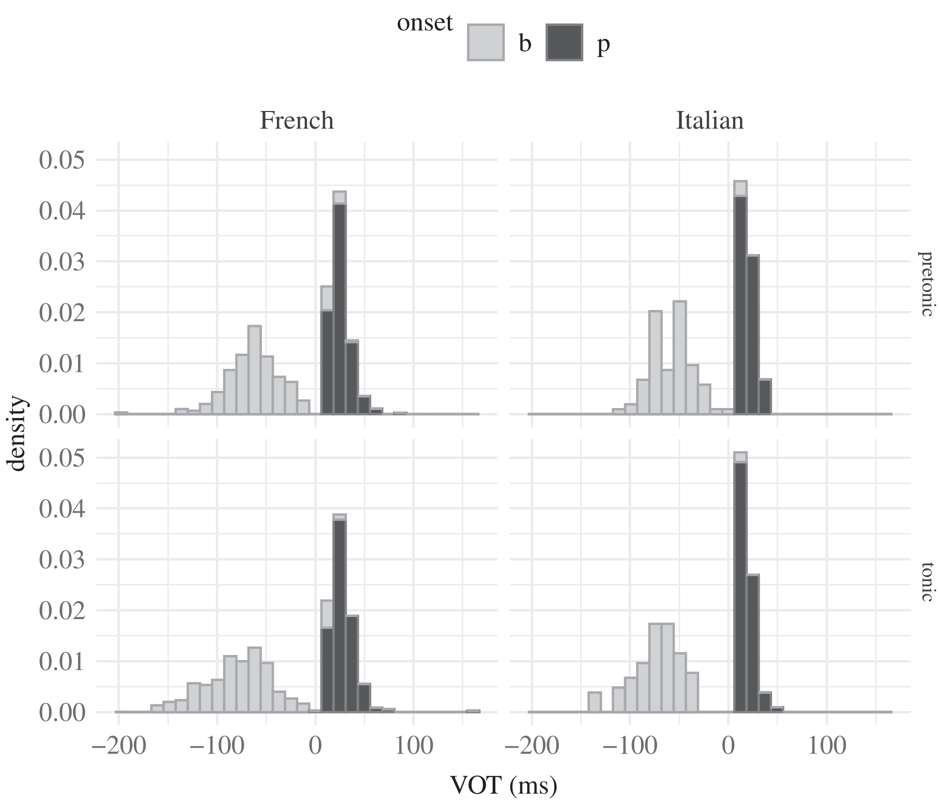

Fig. 2. VOT distributions by voicing, syllable position, and language.

Table 2. Means and standard deviations (msec) for VOT by language, onset, and syllable position.

French				Italian			
Onset	Syllable	Mean	SD	Onset	Syllable	Mean	SD
/b/	Pretonic	–56	34	/b/	Pretonic	–55	24
	Tonic	–69	39		Tonic	–71	27
/p/	Pretonic	26	10	/p/	Pretonic	20	7
	Tonic	28	13		Tonic	18	7

Italian $(\mathbb{E}(\mu_{\text{pretonic}\,/b/}) - \mathbb{E}(\mu_{\text{tonic}\,/b/}) = 16, \text{CI} = [8, 24], P(\delta > 0) = 1)$. Spontaneous devoicing of phonologically voiced plosives occurred for 41 of 486 total tokens (8%) in French and 6 out of 168 tokens (4%) in Italian.

4. Discussion

We find no compelling evidence that a difference in prosodic context (phrase-medial vs in isolation) or syllable position (tonic vs pretonic) influences the F0 following voiced and voiceless obstruents in either French or Italian. In both languages, relative to /m/, we observe F0 raising following the release of /p/ and lowering during the closure of /b/ but not following its release. Given that our Italian model found compelling evidence that F0 following /p/ is greater than /m/ by just 15 Hz, we can be confident that if /b/ had an effect of at least this size, we should be been able to detect it. From this, we infer that the lowered F0 of voiced plosives, relative to sonorants, observed in other languages is unlikely to be driven entirely by the use of an isolated word list context.

One outstanding possibility, which we did not check here (because we focused on keeping the word list as similar as possible in both languages), is that the F0 distinctions in Coetzee *et al.* (2018) were hyperarticulated due to the word list being made up of minimal pairs differing in onset consonant (Baese-Berk and Goldrick, 2009; Kharlamov, 2014). However, we note that the extensive study of Malagasy by Howe (2017) was not based around minimal pairs, and she also finds that F0 is lower following phonologically voiced plosives than when following sonorants, regardless of the presence or absence of closure voicing. Similarly, we might wonder about a difference conditioned by the syllabic structure of the materials. The items analyzed by Coetzee *et al.* (2018) were primarily monosyllabic, and the combination of list intonation and a monosyllable carrier might be more likely to induce an effect of tonal crowding (Arvaniti and Ladd, 2009). However, this possibility strikes us as remote, given that many of the French items in the current study were also monosyllables.

We are left to conclude that F0 following phonologically [+voice] plosives in Afrikaans and Central Malagasy involves a "controlled" low pitch target. However it came about, it cannot simply be a phonologization of the CF0 effect after *voiceless* plosives, and it is unexpected given that the duration of prevoicing does not seem to be a frequent target of enhancement (Mitterer and Ernestus, 2008; but cf. Schertz, 2013). A more plausible explanation may be that offered by Kingston and Diehl (1994), who suggest that F0 is lowered as a listener-oriented strategy to enhance the overall perceptual salience of the [±voice] contrast. In this view, the F0 lowering observed in Afrikaans and Malagasy would follow from the fact that enhancements are learned and language-specific, and need not follow automatically from the presence of universal phonetic contingencies (see also Keyser and Stevens, 2006). An outstanding challenge for phonetic theories is to identify if there exist predictable environments under which such enhancements are more likely (or more likely to be phonologized: Hall *et al.*, 2018), or whether individual differences are ultimately more decisive (Clayards, 2018a, 2018b).

Acknowledgments

Thanks to Brisa Speier-Brito for annotation of the recordings and to two anonymous reviewers for their helpful comments on an earlier version of this manuscript. This research was funded by ERC Grant No. 758605 to J.K. Concept: J.K., D.R.L.; design: D.R.L., Z.E., J.G., J.K.; execution: Z.E., J.G.; analysis: J.G., J.K.

References and links

[1]As a reviewer points out, the existence of a reliable, natural difference in F0 following voiced and voiceless obstruents may already make lowered F0 a candidate for *perceptual* enhancement, which might subsequently be amplified in production. Here, we focus on the question of whether the magnitude of the dichotomy can be reliably modulated by context.

[2]See supplementary material at https://doi.org/10.1121/10.0001698 for online appendixes, and https://doi.org/10.17605/OSF.IO/GX2WH for audio data, EMU database files, and code to reproduce the plots and figures in this paper.
[3]The full list of French items was slightly longer, as we also recorded some additional items not analyzed here.

Arvaniti, A., and Ladd, D. R. (**2009**). "Greek wh-questions and the phonology of intonation," Phonology **26**, 43–74.

Baese-Berk, M., and Goldrick, M. (**2009**). "Mechanisms of interaction in speech production," Lang. Cognit. Process. **24**(4), 527–554.

Bombien, L., Winkelmann, R., and Scheffers, M. (**2018**). *wrassp: An R Arapper to the ASSP Library*, https://github.com/IPS-LMU/wrassp, R package version 0.1.8.

Bürkner, P.-C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Statis. Softw. **80**(1), 1–28.

Clayards, M. (**2018a**). "Differences in cue weights for speech perception are correlated for individuals within and across contrasts," J. Acoust. Soc. Am. **144**(3), EL172–EL177.

Clayards, M. (**2018b**). "Individual talker and token covariation in the production of multiple cues to stop voicing," Phonetica **75**(1), 1–23.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., and Wissing, D. (**2018**). "Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis," J. Phonet. **66**, 185–216.

Coetzee, A. W. (**2019**) (private communication).

Draxler, C., and Jänsch, K. (**2004**). "SpeechRecorder—A universal platform independent multi-channel audio recording software," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, pp. 559–562.

Gao, J., and Arai, T. (**2019**). "Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery," J. Phonet. **77**, 100932.

Hall, K. C., Hume, E., Jaeger, T. F., and Wedel, A. (**2018**). "The role of predictability in shaping phonological patterns," Linguist. Vanguard **77**(s2), 20170027.

Hanson, H. M. (**2009**). "Effects of obstruent consonants on fundamental frequency at vowel onset in English," J. Acoust. Soc. Am. **125**(1), 425–441.

Hombert, J.-M. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic Press, New York), pp. 77–111.

Howe, P. (**2017**). "Tonogenesis in central dialects of Malagasy: Acoustic and perceptual evidence with implications for synchronic mechanisms of sound change," Ph.D. thesis, Rice University, Houston, TX.

Keyser, S. J., and Stevens, K. N. (**2006**). "Enhancement and overlap in the speech chain," Language **82**(1), 33–63.

Kharlamov, V. (**2014**). "Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences," J. Phonet. **43**, 47–56.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Kirby, J. (**2018**). "Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages," J. Phonet. **71**, 326–354.

Kirby, J., and Ladd, D. R. (**2016**). "Effects of obstruent voicing on vowel f0: Evidence from 'true voicing' languages," J. Acoust. Soc. Am. **140**(4), 2400–2411.

Löfqvist, A., Baer, T., McGarr, N., and Story, R. S. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.

Mitterer, H., and Ernestus, M. (**2008**). "The link between speech perception and production is phonological and abstract: Evidence from the shadowing task," Cognition **109**(1), 168–173.

Mohr, B. (**1971**). "Intrinsic variation in the speech signal," Phonetica **23**(2), 65–93.

Schertz, J. (**2013**). "Exaggeration of featural contrasts in clarifications of misheard speech in English," J. Phonetics **41**(3), 249–263.

Winkelmann, R., Harrington, J., and Jänsch, K. (**2017**). "EMU-SDMS: Advanced speech database management and analysis in R," Comput. Speech Lang. **45**, 392–410.