LARYNGEAL CONTRAST AND SOUND CHANGE: THE PRODUCTION AND PERCEPTION OF PLOSIVE VOICING AND CO-INTRINSIC PITCH

Jiayin Gao
Laboratoire de Phonétique et Phonologie (CNRS & Sorbonne Nouvelle)

James Kirby
Institute for Phonetics and Speech Processing, Ludwig Maximilian University of Munich

Inspired by Beddor 2009, this article explores whether and how trading relations between coarticulatory source and effect may serve as a precursor for sound change. It aims at extending the case of vowel nasalization examined by Beddor to the relationship between closure voicing (source) and co-intrinsic pitch (effect). Through four production and perception studies, we show that the inverse source-effect relation observed for vowel nasalization is not found in the voicing contrast of French, a true-voicing language. Instead, we propose that the phonologization of co-intrinsic pitch (a.k.a. tonogenesis) originates from spontaneous devoicing (a production bias), which subsequently triggers an upweighting of pitch (a perceptual adaptation strategy).*

Keywords: trading relation, voicing, co-intrinsic pitch, phonologization, tonogenesis, sound change, French

1. Introduction.

1.1. Co-intrinsic pitch perturbations and sound change. This article aims to develop our understanding of the phonetic origins of sound change (Beddor 2023) through a study of two cues to the laryngeal (voicing) contrast: closure voicing and pitch. At least since House & Fairbanks 1953, it has been observed that differences in consonant voicing are associated with coarticulatory differences in vowel fundamental frequency (F0) in a wide range of languages (for crosslinguistic comparisons in typologically diverse languages, see e.g. Bradshaw 1999, Chen & Downing 2011, Dmitrieva et al. 2015, Hombert et al. 1979, Kirby 2018, Lee 2008, Shimizu 1990, and Tang 2008). The most robust finding is that vowel F0 is raised, relative to the presumed intonational baseline, following phonologically voiceless obstruents (Hanson 2009, Kirby & Ladd 2016), but lowered F0 following phonologically voiced obstruents has also been observed (Coetzee et al. 2018, Howe 2017). In this article, we refer to this effect as co-intrinsic pitch or CF0 (Kingston 2007, Krug et al. 2021), part of the more general phenomenon of microprosody or micromelody (Di Cristo & Hirst 1986, Kohler 1982). While the magnitude, direction, and temporal extent of CF0 perturbations are heavily contingent on the (local and global) prosodic context (Chen et al. 2023, Chen 2011, Choi et al. 2020, Hanson 2009, Jun 1993, Kirby & Ladd 2016, Kohler 1982, Silverman 1986, Xu & Xu 2021), CF0 has been shown to affect the perception of laryngeal contrasts (Abramson & Lisker 1985, Haggard et al. 1970, Whalen et al. 1990).

* This project has been funded by ERC Grant 758605 ‘The emergence and evolution of linguistic tone’ to James Kirby. Many thanks to Marc Brunelle, Dan Dediu, audiences at PaPE 2021 (Barcelona), SRPP (Paris), the University of Edinburgh, and the Institute for Phonetics and Speech Processing at LMU Munich for comments on earlier portions of this work. The authors are profoundly grateful to Bob Ladd for inspiring discussions at multiple stages of this project, as well as to Pam Beddor; editors Andries Coetzee, Meredith Tamminga, and Shelome Gooden; and two anonymous referees for insightful comments and suggestions. The authors alone remain responsible for any errors of fact or interpretation. All experiments in this study have been approved by the ethics committee of PPLS (No. 34-1920/3), University of Edinburgh. Online supplementary materials for this article, including appendices, stimuli files, KlattGrid scripts for stimuli creation, data sets, R scripts for analyzing data, and main results in R markdown output files, are hosted on Project MUSE at http://muse.jhu.edu/resolve/234, and can also be found on the OSF platform at https://doi.org/10.17605/OSF.IO/PY9NR.

Crucially, it is widely assumed both that CF0 plays a central role in the emergence and evolution of lexical tone and that phonologization of CF0 is an essential step in this process (Dockum 2019, Gehrmann 2022, Haudricourt 1961, Hyman 1976, Hyslop 2022, Kingston 2011, Li 1980, Maspero 1912, Matisoff 1973, Mazaudon 2012b, Michaud & Sands 2020, Pittayaporn 2009, Ratliff 2015, inter alia).

An outstanding theoretical and experimental challenge in work on sound change has been to spell out the conditions under which synchronic coarticulatory effects like CF0, nasalization, or vowel-to-vowel coarticulation are PHONOLOGIZED, leading to the development of new phonological contrasts such as lexical tone, contrastive vowel nasality, or umlaut (Blevins 2004, Hagège & Haudricourt 1978, Harrington et al. 2019, Hyman 2013). Two leading models that have addressed this challenge are those of Ohala (1989, 1993) and Beddor (2009, 2012, 2023). In Ohala's model, coarticulation rarely leads to sound change, because adult listeners are sensitive to induced variation and compensate for it in speech perception (Mann 1980, Mann & Repp 1981, Pardo & Fowler 1997). As a result, even when presented with a coarticulated acoustic signal, listeners are typically successful at mapping it onto the phonological category intended by the speaker. It is only in the rare instances when listeners fail to compensate for coarticulation that a sound change will take place. In Ohala's model, sound change is fundamentally abrupt, the result of a perceptual parsing error on the part of the listener (an idea with antecedents going back at least to de Courtenay 1972 [1871] and Passy 1890).

While Beddor's model bears some superficial similarities to Ohala's, it differs crucially in two respects. First, the source of change is not listener error, but parsing variability: Beddor proposes that the flexibility observed in cue weighting (Francis et al. 2000, Harmon et al. 2019, Holt & Lotto 2006, Idemaru & Holt 2011, 2014, Lisker 1986) may lead to individual differences in cue weights (Clayards 2018a, Idemaru et al. 2012, Yu 2021). Second, in Beddor's model, coarticulation is phonologized in an incremental fashion, as the relationship between coarticulatory source and effect changes. Specifically, Beddor proposes that listeners may treat covarying cues to a phonological contrast as PERCEPTUALLY EQUIVALENT (as defined in Best et al. 1981). Over time, one cue-weighting pattern can become dominant, giving rise to sound change. Thus, instead of listener parsing errors, it is the articulatory covariation and perceptual equivalence between cues that provide the seed of sound change.

In support of her model, Beddor (2009, 2012) demonstrates the existence of variable perceptual strategies for identifying nasalization in vowel-nasal (ṼN) sequences in American English: some listeners base their judgments of whether a lexical item contains a nasal (e.g. *bent* vs. *bet*) primarily on information in the N, while others make greater use of information in the Ṽ, or some combination of the two. She further provides acoustic evidence that the durations of Ṽ and N are negatively correlated within and across contexts and speakers, regardless of whether the nasal consonant is robustly present or reduced. Beddor argues that these converging articulatory and perceptual findings provide a plausible basis for the reinterpretation of coarticulatory source and effect and subsequent diachronic change.

This model, which we refer to as the COARTICULATORY PATH MODEL (or CoPath), provides a principled framework in which to consider many types of sound change thought to arise from the phonologization of coarticulatory effects, as well as a methodology for investigating them. Indeed, Beddor herself suggests that 'the expectation is that articulatory covariation and perceptual equivalence might hold for other coarticulatory sources and their effects. They may therefore underlie other sound changes in which the source is lost and the coarticulated variant becomes distinctive' (2009:817). Beddor goes on to suggest that the phonologization of co-intrinsic pitch may be just such a case, based on studies showing the perceptual trading relation between VOT (voice onset time, an index of the temporal interval between the onset of vocal-fold vibration and the release of a plosive: Lisker & Abramson 1964) and F0 of the following vowel (Whalen et al. 1993). Beddor's model has subsequently been cited as providing a plausible mechanism to explain so-called 'tonogenetic' sound changes involving phonologization of CF0 (Coetzee et al. 2018, Howe 2017, Kuang & Cui 2018, Yang et al. 2015).

However, there are reasons to doubt that either articulatory covariance or perceptual equivalence obtain in the case of voicing and CF0, because the relationship between coarticulatory source and effect in this case is considerably more complex than in nasalization. In VN sequences, the velum-lowering gesture is timed to coordinate with the oral closure, producing the acoustic effect of a nasal obstruent. As explained in §1.2 below, if this gesture occurs prior to oral closure, the vowel will be realized as (at least partially) nasalized. We may thus regard the N as the coarticulatory source, and the nasalized Ṽ as the coarticulatory effect, with a single gesture (velum lowering) connecting the two. Because source and effect are mediated by a single gesture, they are treated as equivalent by speakers and listeners. That is, as the source wanes, the effect becomes stronger, and vice versa.

In the case of CF0, this relationship is not quite as clear-cut. First, as we discuss in more detail in §1.3 below, there are probably at least two CF0 effects—raised F0 following voiceless obstruents, and lowered F0 during and/or following voiced obstruents—that come about due to very different laryngeal adjustments. It is thus not obvious what we should regard as the 'coarticulatory effect' in this case—F0 raising, or F0 lowering? Moreover, the production of voicing (or voicelessness) differs from the production of nasalization in that more than one gesture is involved. Voicing (that is, sustained vocal-fold oscillation during oral closure) is the result of a complex coordination of gestures aimed at adducting the vocal folds while also achieving a precise combination of airflow and vocal-fold tension necessary to sustain vibration; similarly, voicelessness requires both abduction of the vocal folds (via displacement of the arytenoid cartilages, an action of the posterior cricoarytenoid muscles) in concert with sufficient vocal-fold stiffness (controlled mainly by actions of the cricothyroid and thyroarytenoid muscles: Titze 2000). In both cases, the acoustic effect is clearly not the result of any one single gesture. The final and perhaps critical point of difference is that the relationship between these constellations of laryngeal gestures and the primary acoustic characteristic of the coarticulatory source is nonlinear. There exist multiple combinations of transglottal pressure differentials and glottal widths under which vocal-fold oscillation is possible for a given stiffness of the vocal folds (Halle & Stevens 1971). Thus, in the case of voicing and CF0, it is not clear that any gestural basis exists for expecting there to be an inverse relationship between the strength of the coarticulatory source (voicelessness or voicing) and the prominence of the coarticulatory effect (raised or lowered F0).

To get a better understanding of how CF0 comes to be phonologized, we first need greater empirical clarity on the relationship between voicing and CF0. To this end, the present study considers whether articulatory covariation and perceptual equivalence obtain between voicing and CF0 in French, a 'true-voicing' language,[1] that is, a language

[1] For more details on the dichotomy between 'true voicing' and 'aspirating', see Beckman et al. 2013, Honeybone 2005, and Iverson & Salmons 1995; for problematic cases, see, for example, Gao & Arai 2019.

in which phonologically voiced obstruents are canonically realized with robust closure voicing. The reason it is critical to study this coarticulatory relationship in a true-voicing language is that in most documented cases in which CF0 has arguably been phonologized, such as Western Kmhmu' (Premsrirat 2001, Svantesson & House 2006), Central Malagasy (Howe 2017), or Afrikaans (Coetzee et al. 2018), evidence suggests that the laryngeal contrast had previously been of the true-voicing type. We have no reason to suspect that CF0 is phonologizing in French, but this is precisely why it is a useful test case: if we are able to observe covariation in production and/or perceptual equivalence between closure voicing and CF0 in a language where the phonological contrast appears to be stable, this would provide strong evidence for the existence of a putative phonetic universal and permit us to make inferences about how the process may have taken place in other languages. In any case, the findings will help us to better understand the phonetic underpinnings of a common sound change.

Our study has two goals. The first is to extend Beddor's (2009) methodology to study the relationship between closure voicing and CF0 in both production and perception in French. We emphasize that our study is not designed as a test of the CoPath model per se, but rather of whether the empirical underpinnings that hold for the case of coarticulatory vowel nasalization also hold for the relationship between voicing and CF0. The second goal is to consider the implications of our findings for the phonologization of CF0 and for coarticulatory sound change more generally. In the remainder of this introduction, we review the essential components of the CoPath model based on the case of vowel nasalization, before laying out the theoretical and methodological bases for our investigation into the case of voicing and CF0.

1.2. ARTICULATORY COVARIATION AND PERCEPTUAL EQUIVALENCE: THE CASE OF VOWEL NASALIZATION. The CoPath model postulates that a robust trading relation—perceptual equivalence and articulatory covariation between coarticulatory source and effect—prefigures at least some instances of phonologization. Such a relation is characterized by an INVERSE correlation between source and effect: when the source is weaker, the effect is stronger, and vice versa. The basis for this assumption is that at least some coarticulatory sources and effects can be traced to a single gesture, and because of this shared gestural basis, listeners may treat the source and its effect as equivalent.

Beddor (2009) presents evidence consistent with this hypothesis from a study on vowel nasalization in American English, where the vowel in /VN/ sequences is typically nasalized to some degree—more so when the /N/ is followed by a voiceless obstruent than by a voiced one (e.g. in *bent* vs. *bend*). Beddor proposes an underlying mechanism that could account for covariation between vowel and consonantal nasality: the variable alignment of a roughly constant-sized nasal gesture with the oral articulators. She hypothesizes that the velum-lowering gesture is approximately stable in terms of temporal extent and magnitude, but the overlap between the nasal and oral gestures may be tighter or looser due to their variable alignment, resulting in a greater or lesser degree of nasality in the preceding vowel. That is, when physiological and acoustic-auditory constraints give rise to an incompatibility between nasality and a following voiceless obstruent, the nasal gesture is considerably anticipated, reducing the nasal consonant and leading to longer (and possibly stronger) vowel nasalization, as schematized in Figure 1. Because the total duration of nasality is relatively stable, when the nasal consonant is shortened, the nasal duration in the vowel becomes longer. Accordingly, vowel nasality is longer and consonantal nasality is shorter in $\mathrm{VNC}_{\mathrm{voiceless}}$ than in $\mathrm{VNC}_{\mathrm{voiced}}$, and the (acoustic) durations of vowel nasality and consonantal nasality correlate negatively.

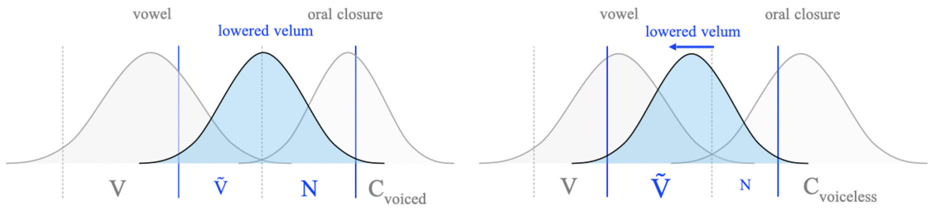

FIGURE 1. Oral-nasal gestural alignment in two consonantal contexts, adapted from Beddor 2009. Lines indicate acoustic representations: acoustic signals corresponding to the vowel part are within dashed (gray) lines; acoustic signals corresponding to the nasal part are within solid (blue) lines. Left panel for weaker/shorter vowel nasalization (smaller-sized letter) and stronger/longer consonantal nasalization (bigger-sized letter); right panel for the inverse relationship of nasality.[2]

On the perception side, Beddor finds evidence for perceptual equivalence between coarticulatory source and effect. Listeners are sensitive to the total duration of nasality, but not to whether nasality is present in the consonant or the preceding vowel. In other words, listeners 'do not perceptually differentiate source from effect' (Beddor 2009:815). In her study, listeners treat source and effect as equivalent across a wide range of stimulus conditions, EVEN WITHIN AN UNAMBIGUOUS CATEGORY. It is this finding that is critical for establishing within-category perceptual equivalence (see also Best et al. 1981, Fitch et al. 1980), a qualitatively stronger effect than cue trading at the category boundary (Repp 1982, 1983).

Recent findings suggest that the strongly negative correlation between source and effect of nasality might be specific to American English. In Standard German, Carignan et al. (2021) do not find a strictly inverse relation between the temporal extent of vowel nasality and consonantal nasality. Instead, the reduction of the coarticulatory source—when the nasal consonant is followed by a voiceless obstruent—has a much more attenuated effect in German compared to English. On the basis of this crosslinguistic difference, Carignan et al. propose that trading relations develop in two stages. In the first stage, represented by German, the increase in the Ṽ portion is much smaller than the decrease in the N portion. However, while the degrees of change in the source and the effect are different, this still allows for an increase in the informativeness of the vowel nasality, upweighting listeners' perceptual use of this coarticulatory effect. This may lead to the second stage, represented by American English, where a source-effect covariation is generalized in both production and perception. Despite the differences in the extent of source-effect covariation reported in the two studies and their implications for the stages of sound change, in both Beddor's (2009) and Carignan et al.'s (2021) proposals, the reduction of coarticulatory source and the perceptual trading between source and effect are considered preconditions for cue reweighting.

1.3. VOICING AND CO-INTRINSIC PITCH: GESTURAL AND PERCEPTUAL CHARACTERISTICS. As detailed above in §1.2, Beddor (2009) proposes that the gestural basis for the source-effect covariation in vowel nasalization is a single nasal gesture variably aligned with oral articulators. While a similarly variable coordination between the laryngeal and supralaryngeal gestures involved in the production of plosive voicing is theoretically

[2] This figure, like several others in the article, is presented in color in the electronic versions of this article, but in grayscale in the print version. Color versions of the figures are also available open access, along with the other supplementary materials, at http://muse.jhu.edu/resolve/234.

possible, there are some fundamental differences in the gestural characteristics between nasal and laryngeal articulations, from which may follow rather different acoustic and perceptual consequences.

As mentioned in §1.1, the basic acoustic relationship between voicing and CF0 is fairly well established: generally speaking, pitch is raised following phonologically voiceless consonants and, at least in some cases, lowered following phonologically voiced consonants.[3] However, what ultimately gives rise to this dichotomy has been the source of much debate. The raised F0 following the release of voiceless plosives has been proposed to result from gestures implemented to suppress vocal-fold vibration during the oral occlusion (Halle & Stevens 1971, Hanson 2009, Kirby & Ladd 2016, Löfqvist et al. 1989). By contrast, the lowered F0 observed during (e.g. Kirby & Ladd 2016) and/or following (e.g. Coetzee et al. 2018) voiced obstruents has been proposed to result from larynx lowering to support vocal-fold vibration during the oral occlusion phase (Honda et al. 1999, Hoole & Honda 2011, Reinholt Petersen 1983, Solé 2018) or a deliberate enhancement of the voicing feature (Kingston & Diehl 1994). Aerodynamic factors have also been argued to play a role (Chi et al. 2019, Francis et al. 2006, Guo & Kwon 2022, Hombert 1976, Kim et al. 2018, Kohler 1982, Ladefoged 1973, Luo 2018, Xu & Xu 2003). As all of these factors probably contribute to the CF0 effect to at least some degree (Hoole & Honda 2011), it does not make conceptual sense to think of a single 'laryngeal gesture', or even gestural complex, as on par with something like the movement of the velum.

In addition, there is evidence suggesting that the coordination between laryngeal and oral articulators is temporally constrained (Benguerel et al. 1978, Moisik et al. 2021, Ohde 1984). For example, in the production of certain types of plosives, peak glottal opening is found to be coordinated with oral release (Kingston 1990). While the precise details of how laryngeal-oral coordination is constrained across consonant types, syllable structures, languages, and speakers remains debated (Goldstein 1990, Hoole 2006), the temporal alignment between laryngeal and oral gestures in a given language may not be as variable as what is found for the velum-lowering gesture. Finally, unlike nasality, with its consistent acoustic fingerprint, variability in laryngeal-oral coordination may have very different acoustic-auditory consequences, including voice murmur, burst, pre- or post-aspiration, breathiness, glottalization, and pitch differences. Therefore, it remains unclear how these acoustic characteristics covary and to what degree they are treated as perceptually equivalent.

In sum, the production of a laryngeal contrast involves coordination between multiple articulators and a complex mapping between multiple articulators and acoustic cues. To trace the pathway toward the phonologization of the coarticulatory effect of a laryngeal articulation, a crucial question is whether the primary cue (voicing) and a secondary cue (CF0) exhibit covariation in production and equivalence in perception before or at the initial stages of phonologization, as observed for vowel nasalization. Below we first discuss the lack of a single shared gestural basis in the case of voicing and co-intrinsic pitch. We then discuss two theoretical accounts that may predict acoustic covariation and perceptual equivalence between voicing and CF0.

Aerodynamic constraints in the production of voicing and CF0. It is frequently observed that voiced obstruents are spontaneously devoiced, that is, phonetically

[3] More robustly, pitch seems to be lowered during the production of phonetically voiced obstruents, which strictly speaking is an intrinsic rather than co-intrinsic effect (Krug et al. 2021). Furthermore, in some languages, the dichotomy is not always apparent; see, for example, Shi et al. 2020 and Gordon 2016.

realized without glottal pulsing during the oral closure phase. Crosslinguistically, the rate of devoicing ranges from sporadic to frequent (Dmitrieva et al. 2015, Hutin et al. 2021, Keating et al. 1983, Kirby & Tan 2023, Pape & Jesus 2015, Pinget et al. 2020, Takada et al. 2015, inter alia). While prosodic and positional factors have been shown to play a role (Davidson 2016, DiCanio & Sharp 2020, Gao & Arai 2019, Hutin et al. 2021), devoicing is also physiologically motivated, largely due to what is best known as the AERODYNAMIC VOICING CONSTRAINT (AVC), which observes that phonation is difficult both to initiate and to sustain throughout the closure or constriction of an obstruent (Ohala 1983, 2011). As the oral pressure increases with the accumulation of air in the oral cavity, the subglottal-oral air pressure difference may fall below the threshold required to sustain glottal vibration. From this it follows that the earlier voicing is initiated, the more challenging it may be to sustain throughout the closure. To overcome the AVC, speakers may deploy a number of synergistic gestures, including prenasalization, pharyngeal expansion, and larynx lowering, in order to expand the volume of the supralaryngeal cavity to facilitate vocal-fold vibration (Ahn 2018, Bell-Berti 1975, Honda et al. 1999, Hoole & Honda 2011, Kingston & Diehl 1994, Solé 2018, Westbury 1983, Zhang & Goldstein 2023). If a speaker successfully manages to circumvent the AVC by means of one or more of these maneuvers, the magnitude of the coarticulatory effect of that maneuver on the adjacent vowel would be expected to be greater than if the speaker had not implemented that maneuver, all else being equal. As one of these effects, the co-intrinsic lowering of F0 would thus be expected to correlate POSITIVELY, rather than inversely, with the degree of closure voicing. In other words, this relation would be the opposite of the source-effect relation observed for nasalization.

COMPENSATORY CUE ENHANCEMENT. Nonetheless, even if multiple cues signaling a contrast are not related by a shared articulatory basis, reduction of a primary cue could still lead to an increase in one or more secondary cues as an attempt to enhance an acoustic/auditory target or a phonological contrast (Carignan et al. 2011, Ferlus 1979, Gao et al. 2021, Gay et al. 1981, Keyser & Stevens 2006, Kingston & Diehl 1994, Kühnert et al. 1991, Liljencrants & Lindblom 1972, Maeda 1990, Mazaudon & Michaud 2008, Perkell et al. 1993, Perrier & Fuchs 2015). In the case of prevoiced obstruents, this would lead to the expectation that as the duration of vocal-fold vibration during the closure decreases, a secondary cue such as CF0 may be produced with greater magnitude and/or temporal extent; as a result, CF0 would correlate inversely with closure voicing. Such compensatory enhancement has been hypothesized by some authors to be gradient and probabilistic, predicting covariation between cues across categories as well as within a single phonological category (Clayards 2018b, Kirby 2013, Shultz et al. 2012).

Despite such theoretical predictions, the empirical grounding for an inverse relationship between voicing and onset F0 is lacking. In the case of voicing contrast, while there is some evidence for covariance of multiple cues across voicing categories (Clayards 2018b, Shultz et al. 2012), studies in several languages (English, Spanish, French, and Japanese) have so far failed to find consistent evidence of inverse covariance between VOT and onset F0 WITHIN the voiced or voiceless category (Clayards 2018b, Dmitrieva et al. 2015, Gao & Arai 2019, Kirby & Ladd 2015).

AUDITORY-BASED PERCEPTUAL EQUIVALENCE BETWEEN VOICING AND CF0. If there is no articulatory or compensatory basis for voicing and CF0 to inversely covary, why is there any reason to think they might be treated as perceptually equivalent? One potential source of perceptual equivalence could be auditory cue enhancement. John Kingston, Randy Diehl, and colleagues have made a strong case that auditorily similar cues to voicing—crucially, including closure voicing and CF0—contribute to an intermediate perceptual property (IPP) they term the 'low-frequency property' (Kingston & Diehl 1994, Kingston et al. 2008). An IPP is an integrated perceptual effect of auditorily similar (as opposed to merely covarying) acoustic properties, which contribute to a percept by virtue of being mutually reinforcing. Kingston and colleagues hypothesize that, because different combinations of acoustic properties with various values can signal the same IPP (such as the low-frequency property), the acoustic properties involved will be perceptually equivalent (Kingston et al. 2008:29).

If closure voicing and CF0 cohere into an IPP, we might then expect to find them in a trading relation across the entire stimulus space. Testing this requires a task that examines within-category perceptual equivalence, such as discrimination (Best et al. 1981, Fitch et al. 1980) or fixed classification experiments (Garner 1974), which do not require explicit labeling. Kingston et al. (2008), using several fixed classification experiments, showed that English-speaking listeners perceive low-rising F0 and closure voicing in VCV sequences as equivalent. To the best of our knowledge, however, within-category perceptual equivalence has not been tested for closure voicing and CF0 in a true-voicing language.

In a categorization task, although there is evidence that perceptual trading between VOT and CF0 occurs across the stimulus space in an aspirating language (Whalen et al. 1993), the perceptual role of CF0 in a true-voicing language has been found to be limited in the acoustic space. Llanos et al. (2013) found that for Spanish listeners, CF0 contributed to identification of the voicing contrast only in the voiceless region of the stimulus space, while Gao et al. (2019) and Kang and Hirayama (2023) observed this effect only in the voiced region for Japanese listeners. This is the second reason why it is important to collect more data testing the perceptual relation between voicing and CF0 across the entire stimulus space.

1.4. Aims and protocol of the present study. As reviewed in §1.1 and §1.2, the CoPath model proposes that phonologization of the coarticulatory effect is driven by mutual covariation in production and perception, with an inverse relationship between two cues in production providing the basis for listeners to treat them as perceptually equivalent. Given the success of this model in accounting for the phonologization of coarticulatory nasalization (Beddor 2009, Beddor et al. 2018), it has been suggested that it could also account for other instances of phonologization, such as the emergence of contrastive F0 height from an onset voicing contrast. As discussed in §1.3, however, it is not clear that the conditions that obtain in the case of nasalization also hold in the case of voicing and CF0. Therefore, our primary aim with this study is to assess whether, and to what extent, articulatory-acoustic covariation and/or perceptual equivalence obtain between voicing and CF0 in Metropolitan French, a true-voicing language. In particular, we seek to establish (i) whether voicing covaries with the magnitude of CF0 in production, both within and across categories, and (ii) whether closure voicing and CF0 are treated as perceptually equivalent both within the voiced category and at the voiced-voiceless category boundary—that is, whether they trade across a wide range of stimulus conditions, not just in configurations where the primary cue is ambiguous.

We begin by looking for evidence of within- and across-category covariation in production by considering how onset F0 correlates with the duration of prevoicing. Going forward, we refer to the acoustic manifestation of closure voicing as 'prevoicing' and to the acoustic value of the CF0 effect as 'onset F0' (i.e. F0 at the following vowel onset). Prevoicing is (at least utterance-initially) indexed by VOT, which is negative for a typical voiced plosive in French. In particular, we assess whether VOT is in an inverse relationship to onset F0. We also consider two predictions based on the compensatory-enhancement perspectives outlined in §1.3. First, at least some accounts would predict that VOT-onset F0 covariation should be observed ACROSS the voicing categories as well as WITHIN the voiceless and the voiced categories (e.g. Clayards 2018b). Second, if the degree of compensation is proportional, onset F0 should become lower as the duration of prevoicing decreases, and should be lowest following DEVOICED plosives compared to canonical voiced plosives.

Next, we look for evidence of perceptual equivalence using three experimental paradigms. Building on previous work (Beddor 2009, Best et al. 1981, Fitch et al. 1980), we conduct an AX discrimination (same/different) experiment to test for perceptual equivalence between prevoicing and low onset F0 in a true-voicing language. While a number of previous studies have addressed the question of perceptual integration of VOT with other cues to voicing, such as first-formant cutback (Lisker 1975, Lisker et al. 1977, Repp 1983, Serniclaes 1987, Simon & Fourcin 1978, Soli 1983, Stevens & Klatt 1974), we know of only one study that examines the perceptual integration of onset F0 and (presence vs. absence) of prevoicing by English-speaking listeners (Kingston et al. 2008). In the present study, we compare two cases with French-speaking listeners: (i) when the duration of prevoicing varies but falls clearly within the voiced category, and (ii) when prevoicing is absent from a phonologically voiced plosive (i.e. at the voiced-voiceless boundary). We focus on these cases because, as mentioned above, some versions of auditory-based theories predict that it is the presence of prevoicing—but not its duration—that creates auditory similarities with a low F0 at the following vowel onset, giving rise to a trading relation between the two cues. If true, the absence of prevoicing would create a categorization ambiguity between voiced and voiceless plosives in a true-voicing language (but not an aspirating language), leading to a contrast reduction that could initiate the phonologization of CF0. At least some theories posit that contrast reduction, either prior to or concomitant with enhancement, is the precursor of the phonologization of a secondary cue (e.g. Bang et al. 2018, Haudricourt 1961, Kirby 2013).

We then employ the TWO-ALTERNATIVE FORCED-CHOICE (2AFC) and VISUAL ANALOG SCALING (VAS) paradigms to test a more classic trading relation in word identification involving voicing categorization, paying special attention to the role of onset F0 in ambiguous vs. unambiguous VOT regions. Our decision to deploy VAS alongside the more traditional 2AFC paradigm is based on a growing body of research suggesting that categoricalness in speech perception may be at least partly encouraged by the dichotomous nature of the 2AFC task (McMurray 2022). The findings of VAS studies (Kapnoula et al. 2017, Kim et al. 2020, Kong & Edwards 2016, Massaro & Cohen 1983) show that listeners' responses are more gradient, and their use of secondary cues is greater in the VAS task than the 2AFC task. Thus, the contribution of CF0 to voicing judgments in unambiguous VOT regions may be more apparent in VAS responses than in traditional 2AFC responses.

Our specific research questions can be summarized as follows.

IN PRODUCTION:

RQ1: Does VOT covary with onset F0 WITHIN the voiced and/or voiceless category?

RQ2: Does VOT covary with onset F0 ACROSS the voiced and voiceless categories?

RQ3: Are voiced plosives that are devoiced followed by a lower onset F0 than those that are typically prevoiced?

IN PERCEPTION:
RQ4: Is longer prevoicing perceived as equivalent to a lower onset F0 across a wide range of stimulus conditions, whether VOT is ambiguous or not?
RQ5: Does onset F0 affect voicing categorization along the entire VOT continuum, and not only at the voiced-voiceless boundary?

2. STUDY 1: TESTING FOR COVARIATION IN PRODUCTION BETWEEN VOICING AND CF0. Study 1 examines correlations between VOT and onset F0 both within the voiced and voiceless categories and across the two categories.

2.1. PARTICIPANTS AND MATERIALS. Data for this study were drawn from a previous production experiment (Kirby et al. 2020) in which covariation between voicing and CF0 was not investigated. Eleven native speakers of French (nine women and two men), from eighteen to twenty-three years old (median = 21, mean = 21, *SD* = 1.5), participated in the study. Each speaker was recorded reading seventy-two mono- and disyllabic French lexical items with /p, b, m/ onsets in isolated citation form.[4]

2.2. ANALYSIS AND RESULTS. Figure 2 plots the relation between speaker-normalized VOT and onset F0 (operationalized as the averaged F0 over the first 25% of the interval between the plosive release and vowel offset) across twenty-four /p/-onset and twenty-four /b/-onset lexical items. Within the voiced category (further broken down into prevoiced and devoiced) and voiceless category, VOT has either no correlation or a weakly positive correlation with onset F0 (for prevoiced /b/: $R = 0.13$, $p < 0.005$; for devoiced /b/: $R = 0.24$, $p = 0.06$; for /p/: $R = 0.03$, $p = 0.45$).[5] Specifically for the voiced category, this means that the degree of prevoicing has a weakly positive correlation with F0 lowering. This runs against the inverse relation between coarticulatory source and effect. The same pattern is observed for most individuals (Figure 3).

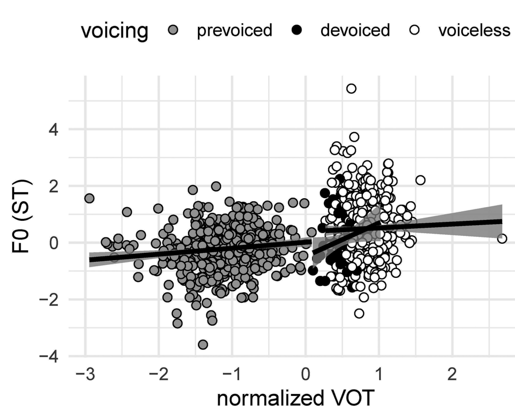

FIGURE 2. Scatterplot of onset F0 against speaker-normalized VOT across all speakers and tokens with linear smoothing.

[4] Details related to the speech materials for the studies discussed in this article, as well as the appendices referenced throughout, can be consulted online at http://muse.jhu.edu/resolve/234 and at https://doi.org/10.1121/10.0001698.

[5] We also used other time intervals for F0 extraction, including the first 30 ms, first 50 ms, and the first 50% of the interval between the plosive release and vowel offset, but none resulted in an inverse relation between prevoicing and onset F0.

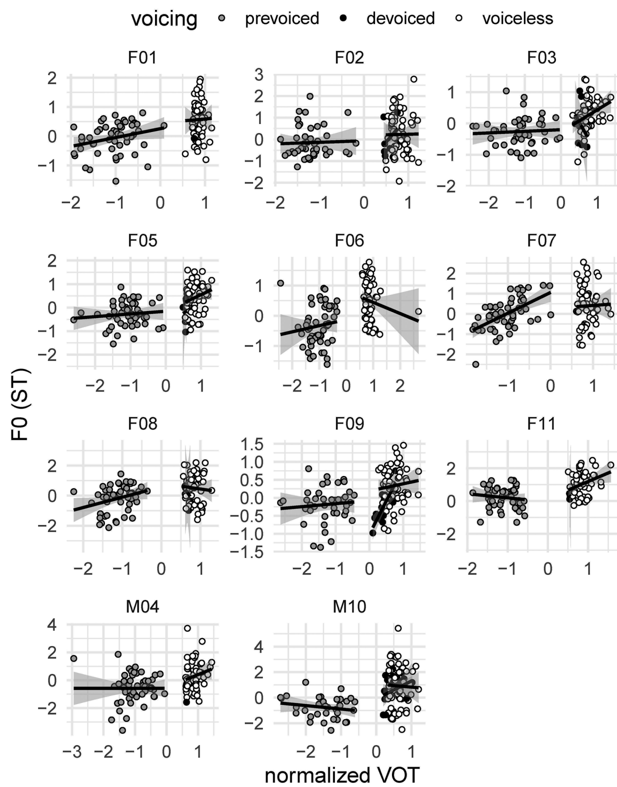

FIGURE 3. Individual scatterplots of onset F0 against VOT.

To further assess the relation between VOT, onset F0, and other factors within and across categories, a linear mixed-effects model was fitted to the onset F0 data (in semitones) using the lmer function in R (R Core Team 2021), with VOICING GROUP (Helmert-coded: prevoiced, devoiced, voiceless), VOT (numerical), SEX (treatment-coded), and the VOT : Voicing group interaction as predictors. Random intercepts were included for participants and items. Online appendix A.1 shows the full model summary, followed by the estimates of the slopes of the regression lines for each level of the Voicing group predictor computed by emtrends (Lenth 2021). ONSET F0 increases with VOT, confirming a weakly positive correlation between the two variables (est. = 0.01, $SE = 0.005$, $t = 2.27$, $p < 0.05$), with no interaction with Voicing group. The regression line is estimated to be flat for the prevoiced category (slope = −0.00006) and positive for the devoiced and voiceless categories (slopes at 0.02 and 0.009, respectively). These results indicate the lack of inverse relationship between onset F0 and VOT not only within each voicing category, but also across the categories.

If speakers were enhancing CF0 in a compensatory manner, they might be more likely to do so in instances where prevoicing fails to be produced (i.e. where confusion with voiceless plosives would be more likely). However, as shown in Fig. 2, the instances of /b/ that are devoiced (10.1% on average across speakers) are produced with similar onset F0 to prevoiced plosives. This suggests that onset F0 is not lowered to compensate for the devoicing of voiced stops. Pairwise comparisons based on a linear mixed-effects model with only Voicing group and Sex as predictors and the same random predictors as the model above suggest no difference in onset F0 between the prevoiced and the devoiced categories (est. = −0.16, *SE* = 0.10, *t* = 1.62, *p* = 0.24).

To sum up, our acoustic analysis of isolated words in French shows that VOT has at best a weakly positive correlation with onset F0 within both voiced and voiceless categories (RQ1), as well as across categories (RQ2), and that devoiced plosives are not produced with a lower onset F0 than prevoiced plosives (RQ3).

3. STUDIES 2–4: TESTING FOR PERCEPTUAL EQUIVALENCE BETWEEN PREVOICING AND ONSET F0. Study 1 found no evidence for an inverse relation between voicing and CF0 in production. However, as noted in §1.3, the lack of an inverse correlation between two cues in production does not rule out the possibility that equivalence obtains between them in perception. We looked for evidence of perceptual equivalence between prevoicing and onset F0 using three experimental paradigms: an AX discrimination task (§3.1), a 2AFC task, and a VAS task (§3.2). All experiments were conducted online using the PsyToolkit platform (Stoet 2010, 2017) between October 2020 and February 2021, during COVID-19 lockdowns.

3.1. STUDY 2: AX DISCRIMINATION. To test for perceptual equivalence between prevoicing and onset F0, we adapted the AX discrimination design of Beddor 2009 to assess how listeners discriminated between stimuli pairs with different manipulations of VOT and/or onset F0.

PARTICIPANTS. We recruited 115 native speakers of Metropolitan French from various mailing lists and participant pools in France. Participation was restricted to those who were eighteen years of age or older, but was otherwise not restricted on the basis of sex/gender, social, linguistic (i.e. mono- or multilingual speakers of French), or neurocognitive background.

MATERIALS. All stimuli were *Ca* syllables (where C was a labial plosive), synthesized using the KlattGrid synthesizer implemented in Praat (Weenink 2009), adapting a procedure by Schertz (2014). Syllables were created with two crossed parameters: (1) four levels of VOT: −100, −60, −40, and 0 ms, and (2) two levels of onset F0: 90 Hz and 130 Hz. The F0 of the turning point was set to 130 Hz. That is, from the vowel onset, F0 was either low (rising from 90 to 130 Hz, crossed with two F0 transition lengths: 40 ms and 100 ms) or flat at 130 Hz. The F0 offset was set to 100 Hz—that is, F0 was falling from the turning point till vowel offset—to ensure a relatively natural F0 declination. All other parameters were kept constant, including the following: burst duration at 5 ms and vowel duration at 245 ms.

Stimuli pairs can be organized according to CONDITION and AMBIGUITY.

(A) Condition. On each trial, participants were presented with a pair of stimuli potentially differing in VOT and/or onset F0, yielding four conditions patterned after those in Beddor 2009:799ff.

(A0) Control: stimuli were identical, sharing both the same VOT (−100, −60, −40, or 0 ms) and onset F0 (90 or 130 Hz).

(A1) VOT-only: stimuli differed only in VOT by 60 ms (−100 vs. −40 ms, or −60 vs. 0 ms). Each stimuli pair shared the same onset F0 (90 or 130 Hz).

(A2) Additive: stimuli differed in VOT and onset F0 in such a way that the cues were expected to reinforce each other to increase the perceived difference. That is, one stimulus had LONGER prevoicing and LOW onset F0—both prototypical of /b/—while the other had SHORTER OR NO prevoicing and FLAT onset F0—both less prototypical of /b/ (e.g. −100 ms VOT with 90 Hz onset F0 vs. −40 ms VOT with 130 Hz onset F0).

(A3) Canceling: stimuli differed in VOT and onset F0, but in such a way that the cues were expected to trade with each other, canceling out the perceived difference. That is, one stimulus had LONGER prevoicing and FLAT onset F0, while the other had SHORTER prevoicing and LOW onset F0—one cue as prototypical of /b/ and the other not, in both cases (e.g. −100 ms VOT with 130 Hz onset F0 vs. −40 ms VOT with 90 Hz onset F0).

Note that since we aimed to ensure that the differences would be perceptible by participants, the VOT/F0 difference within a stimulus pair was chosen to be larger than in a typical AX discrimination task along acoustic continua (see also Kingston et al. 2008).

(B) Ambiguity of prevoicing. To probe for evidence of perceptual equivalence between prevoicing and onset F0, we structured the design to examine whether the weight of the F0 cue differed depending on the ambiguity of the primary cue, VOT. This allows us to see if perceptual equivalence holds regardless of whether stimuli are clearly members of the voiced category, or ambiguous between voiced and voiceless. For the nonidentical conditions, half of the stimuli pairs had prevoicing that was unambiguously within the typical /b/ range (−100 vs. −40 ms). For the other half, one stimulus was clearly in the voiced category, but the other was at the category boundary, that is, was ambiguous (−60 vs. 0 ms). For the control condition (A0), VOT was held constant within the stimuli pair. Thus, in terms of VOT, two-thirds of the stimuli pairs were clearly within the voiced category (sharing a VOT of −100, −60, or −40 ms), while one-third were defined as ambiguous (sharing a VOT of 0 ms). Examples of unambiguous stimuli pairs are given in Figure 4 and ambiguous stimuli pairs in Figure 5.

Stimuli. All VOT and onset F0 levels were crossed, yielding eight stimuli pairs for condition A1, repeated ten times, four stimuli pairs per conditions A2 and A3, repeated twenty times, and sixteen stimuli pairs for condition A0, repeated five times, for a total of 320 trials. Pitch transition was manipulated for independent reasons. As the inclusion

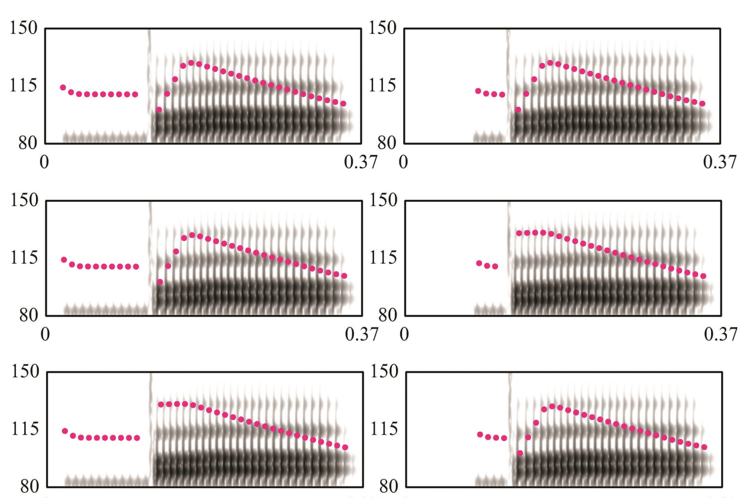

FIGURE 4. Illustrations of the three conditions for stimuli pairs where both are unambiguously /b/ in terms of VOT. Curves represent F0 trajectories. Time (seconds) on *x*-axis and F0 (Hz) on *y*-axis.

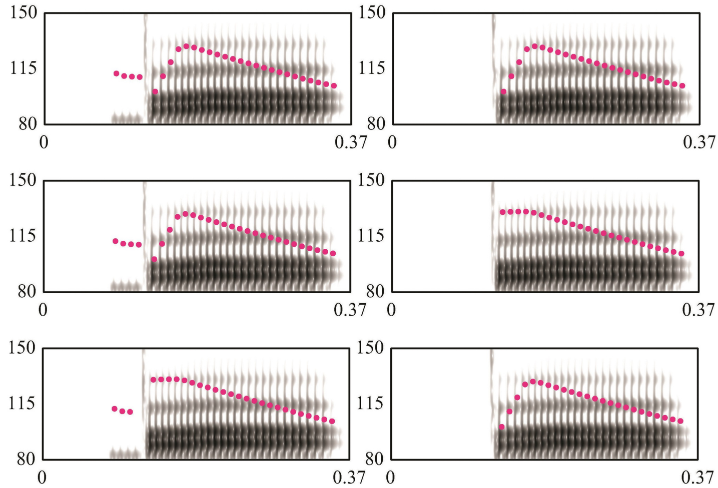

Figure 5. Illustrations of the three conditions where one stimulus is ambiguous between /p/ and /b/ in terms of VOT. Curves represent F0 trajectories. Time (seconds) on *x*-axis and F0 (Hz) on *y*-axis.

of this factor did not change our main findings (see appendices B.2 and C.1), we aggregate across transition lengths for ease of exposition. The experiment also contained an additional condition (appendix B.1) to be reported in future work.

Procedure. Participants were instructed to wear headphones or earbuds and complete the experiment on a computer in a quiet room. All instructions and survey questions were given in French. Prior to completing the discrimination task, participants completed a brief survey on demographic information such as age, sex/gender, place of birth, and so forth.

In the AX discrimination task, participants were presented with two stimuli on each trial and asked to identify whether they sounded identical or different, by pressing the tab or enter key, respectively. The interstimulus interval (ISI) was set to 330 ms, the intertrial interval to 700 ms, and the time-out to 2.5 s. Note that a 330 ms ISI would test an auditory mode of perception (Werker & Logan 1985). The test phase was preceded by a training phase with ten mandatory trials, including four identical stimuli pairs and six different stimuli pairs. Since within-category discrimination is known to be difficult, participants were allowed to complete the training phase twice, with trials repeated if they responded in error. Stimuli in both training and test phases were randomized for each participant.

Following the experiment, participants completed an exit survey that invited them to choose among broad categories related to language, education, and music-related background, and speech/hearing/motor-related disorders. They were invited to give further comments on a voluntary basis. The entire session lasted around twenty to thirty minutes.

Predictions. If prevoicing and low onset F0 are perceived as equivalent across the stimulus space,

(i) discrimination accuracy in the additive condition should be greater than in the VOT-only condition, which should in turn be higher than in the canceling condition (additive > VOT-only > canceling); and
(ii) this pattern should hold regardless of the ambiguity of prevoicing.

In contrast, if listeners discriminate the stimuli pair based on their acoustic (dis)similarity, we expect the discrimination accuracy to follow the order additive = canceling > VOT-only.

Analysis and results. We discarded the data of one participant who failed to press a button (i.e. exceeded response time-out) for 140 of 400 trials of the entire session. Data from 114 listeners (eighty-three women, thirty-one men), from eighteen to sixty-one years old (median = 28, mean = 32, *SD* = 12), were thus retained. Although remote data collection is increasingly accepted as valid even for examining the perception of phonetic details (e.g. Elliott et al. 2022, Sanker 2023), we were aware of its limitations. For example, we could not be certain about the engagement or the experimental environment of the participants. However, given that within-category discrimination is known to be difficult, there was no objective way to discard data based on listeners' performances. To remove possible concerns that the retained data might contain unmeaningful responses, we compared the results of these data with the results after two rejection criteria.[6] Neither criterion changed the main findings we present below.

Discrimination accuracy was indexed by *d′* scores (Macmillan & Creelman 2005). *d′* was calculated for each participant, based on hit vs. miss responses to different stimuli pairs, and false alarm vs. correct rejection responses to identical stimuli pairs. The log-linear method was used to adjust extreme single scores (i.e. all hits or all misses) by adding 0.5 to both the number of hits and the number of false alarms, and 1 to both the number of different-pair trials and the number of identical-pair trials (Stanislaw & Todorov 1999). Figure 6 plots the density of *d′* scores for the three nonidentical conditions by ambiguity. Overall, discrimination accuracy follows the following order: for ambiguous stimuli pairs, additive > canceling > VOT-only; for unambiguous stimuli pairs, canceling = additive > VOT-only.

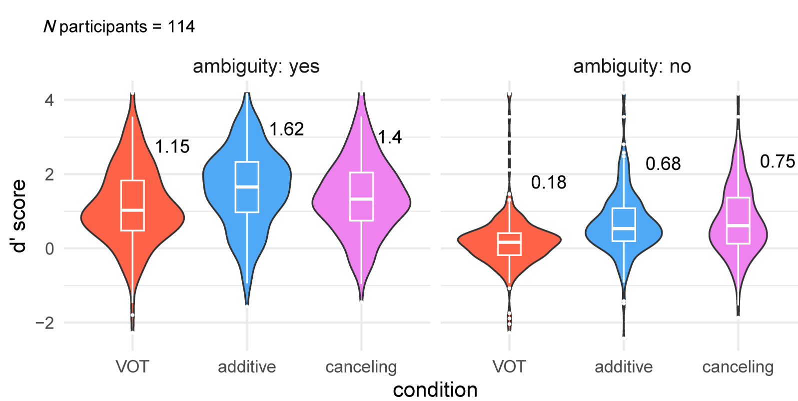

Figure 6. Density plots of *d′* for VOT-only, additive, and canceling conditions by ambiguity. Numbers indicate means.

[6] One criterion was based on listeners' performance during the training phase, and the other on their performance during the VOT-only ambiguous condition, where participants were expected to perform better than the chance level. See appendices B.3.2–B.3.3 for the results of these analyses. In addition, see appendix B.3.4 for the distribution plot of the numbers of timed-out trials, based on which we decided to reject the outlier participant from the retained data.

A binomial generalized linear mixed model (GLMM) was fitted to the response data using the glmer function in R (R Core Team 2021), with Helmert-coded predictors Condition, Ambiguity, and their interaction. Random intercepts were included for participants, and by-participant random slopes for Condition and Ambiguity. The full model summary and the estimated marginal means for each level of the predictors are given in appendix B.3.1. The prediction plot of the model with the interaction term (Figure 7) shows the same pattern as the *d′* plot above. Here we focus on the results that are relevant to the predictions laid out in §3.1.

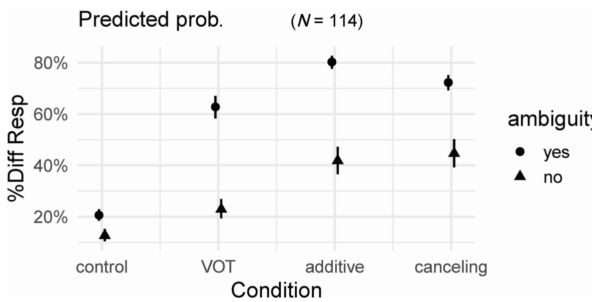

Figure 7. Prediction plot of the GLMM model showing the interaction between Condition and Ambiguity predicting the probabilities of 'different' responses (model: *glmer*(*response* ~ *ambiguity* * *condition* + (1 + *condition* + *ambiguity*)|*subject*)). Error bars indicate 95% confidence intervals.

Pairwise comparisons were conducted using the emmeans package (Lenth 2021). These show that discrimination accuracy (i.e. the probability of giving a 'different' over a 'same' response to nonidentical stimuli pairs) is higher for additive than for canceling conditions only for ambiguous stimuli (est. = 0.45, $SE = 0.06$, $z = 7.1$, $p < 0.0001$), but does not differ between the two conditions for unambiguous stimuli (est. = −0.14, $SE = -0.12$, $z = -1.99$, $p = 0.19$). In both cases, the discrimination accuracy is the lowest for the VOT-only condition (where listeners' probability of a 'different' response was still higher than for the control condition, excluding the possibility of a floor effect).

AX discrimination task: summary. The AX discrimination task was conducted to look for evidence of perceptual equivalence between prevoicing and onset F0. Our results provide evidence for weak perceptual equivalence between prevoicing and low onset F0 for ambiguous stimuli, as indicated by the ranking of discrimination accuracy additive > canceling > VOT-only. But they provide no evidence for unambiguous stimuli, as indicated by the ranking canceling = additive > VOT-only. That is, low onset F0 trades with the absence of prevoicing, but not with its temporal reduction. When prevoicing is completely absent in one member of a stimulus pair (i.e. when VOT = 0), low onset F0 at least partly cancels out the effect induced by the absence of prevoicing. However, this canceling effect appears to be incomplete, because the canceling condition yields a similar discrimination accuracy as for the VOT-only condition. This suggests that while French listeners only partly rely on low onset F0 to substitute for the absence of prevoicing, they are also sensitive to the onset F0 differences without perceiving them as a cue to voicing. Taken together, prevoicing trades perceptually with low onset F0 only when VOT is ambiguous, but a strict perceptual equivalence relationship between the two cues is not established when prevoicing is robust (RQ4).

3.2. Studies 3 and 4: 2AFC and VAS identification.

Participants and procedure. Of the 115 participants who completed the AX discrimination task, seventy-two replied to our invitation to complete the identification tasks. The 2AFC and VAS experiments were also conducted online, at least one month after the AX experiment. In the 2AFC task, on each trial, participants were presented with one stimulus and two response choices, 'pa' and 'ba', and pressed the tab or enter key to indicate whether they thought the sound corresponded to the choice on the left or right side of the screen. Position of the response options was randomized across participants, but constant for each individual participant. The intertrial interval was set to 1.2 s and the time-out to 2.5 s. The test phase was preceded by a short training phase with only six trials, using stimuli randomly chosen from the test phase. No feedback was given during training in order to stress the fact that there were no 'correct' responses and to avoid giving participants the impression that their performance was being evaluated.

After the completion of the short 2AFC task, participants were presented with some animal videos during a short break, before starting the VAS task. In the VAS task, participants were again presented with a single stimulus, but this time, the two response choices were located at the two ends of a horizontal slider, on the same sides as in the 2AFC task for that participant. Participants dragged the vertical bar on the slider from the center point to indicate where the sound stimulus should be located between the left and the right endpoints, that is, between 'pa' and 'ba'. They then pressed the button 'next' to proceed to the following trial. No time-out was set so that listeners could have ample time for each trial. Stimuli were presented in a different randomized order for each participant. In the training phase, participants started with eight trials in a fixed order with synthesized stimuli other than 'pa/ba' to familiarize them with using the slider. They then completed six trials with randomly chosen 'pa/ba' stimuli to familiarize them with the task itself. The whole session lasted around twenty to thirty minutes.

Materials. The primary purpose of the 2AFC task was to determine the extent to which onset F0 influenced voicing category judgments across the VOT continuum. However, we were also interested in inferring how listeners may have been identifying the stimuli previously heard in the AX task. For this reason, we used the same stimuli from the AX task in the 2AFC task, except for two modifications to the VOT levels. First, the −100 ms level was removed because we assumed that all listeners would consistently identify a VOT of −100 ms as voiced. Second, a −20 ms level was added to ensure a linear change along the VOT continuum, following the procedure of most 2AFC tasks.[7] Stimuli varied in VOT (four levels) and Onset F0 (two levels, crossed with two F0 transition lengths). Each stimulus was repeated four times, yielding sixty-four trials. (The 2AFC task could train listeners to give more categorical responses in the following VAS task, which was why we avoided a longer exposure with more trials.)

For the VAS design, stimuli were drawn from a continuum of six levels of VOT × three levels of Onset F0 (crossed with two F0 transitions). VOT ranged from −60 to +20 ms, with a 20 ms step in the negative range and a 10 ms step in the positive range. A modification and an extension were made in the onset F0 levels, with 100, 130, and 169 Hz at the vowel onset, to ensure equal semitone intervals between two consecutive steps (4.5 semitones). Each stimulus was repeated five times, yielding 180 trials in total.

[7] No positive VOT steps were used because we wanted to avoid a change in terms of the VOT distributions in the stimuli between the AX task and the 2AFC task. We were aware that this could encourage listeners to increase their 'pa' responses for stimuli at VOT = 0 ms as the rightmost endpoint of the continuum.

PREDICTIONS. First, as widely shown in previous literature, we predicted that lower onset F0 would lead to more, and more rapid, voiced responses and a shift in the identification curve toward the positive VOT endpoint. If listeners were treating the ambiguous stimuli in the AX task primarily as instances of voiceless plosives, onset F0 should have the greatest influence on voicing identification and judgment at the voiced-voiceless boundary (i.e. VOT at or slightly below 0 ms; see e.g. Serniclaes 1987). By contrast, if onset F0 plays a role across the stimulus space, some effects should be revealed in the VAS task even when VOT is unambiguous.

ANALYSIS. We discarded from all analyses the data of one participant who gave voiceless responses 100% of the time because this was most likely due to a lack of engagement. Data from seventy-one participants (fifty-five women, sixteen men), from eighteen to fifty-nine years old (median = 31, mean = 33, *SD* = 11), were thus retained for analysis. For the 2AFC task, six listeners failed to press a button for more than fourteen of sixty-four trials (i.e. more than 20%), which quite clearly suggested a lack of engagement, and thus their data were further excluded from the plots and the statistical analyses for the 2AFC task.[8]

For 2AFC results, binomial GLMM models were fitted to the binary responses using the glmer function of the lme4 package (Bates et al. 2015). For VAS results, beta regression GLMM models were fitted to the VAS response [0, 100] rescaled to [0.01, 0.99] using the glmmTMB package (Brooks et al. 2017), given the typical beta distribution illustrated by the VAS data found in previous studies (Kapnoula et al. 2017, Kong & Edwards 2016).[9] All post-hoc pairwise comparisons were made using emmeans (Lenth 2021).

RESULTS. Figure 8 shows the aggregated identification rate by VOT and Onset F0 observed in the two identification tasks. Unsurprisingly, in both tasks, the /ba/ identification rate decreases when VOT increases and when onset F0 increases. Of greater interest is to see precisely how onset F0 affects listeners' performance along the entire VOT continuum, which we examine in the following.

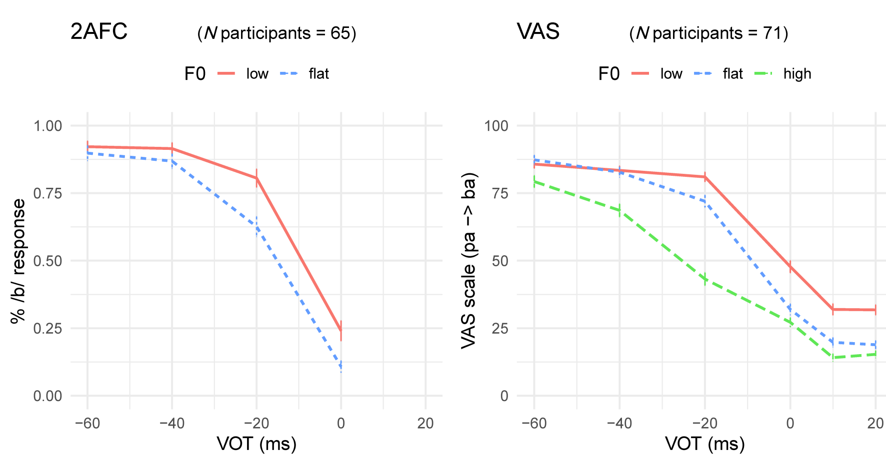

FIGURE 8. Observed identification rate for 2AFC and VAS by VOT and Onset F0. Error bars indicate 95% confidence intervals.

[8] See appendix C.3.1 for the distribution plot of the numbers of timed-out trials. The inclusion or exclusion of these data for the VAS task does not change our main findings. See appendices C.1, C.2, and C.3.3 for results including all participants.

[9] The beta distribution illustrated by our VAS data may be obscured by the peak at the midpoint of the VAS scale (appendix C.4.1). However, this peak is possibly due to the lack of engagement of some participants, who responded excessively with the midpoint of the scale (see individual plots in appendix C.2).

2AFC responses. A binomial GLMM model was fitted to the response data, with the centered and standardized continuous predictors VOT, (Onset) F0, and their interaction. Random intercepts were included for participants, and by-participant random slopes for the interaction between VOT and F0. The full model summary is given in appendix C.3.2. Figure 9 plots the predicted probability of /b/ response by VOT and F0. /b/ response is almost at ceiling when VOT is at −60 and −40 ms, where F0 has little effect. /b/ response is mostly affected by F0 when VOT is at −20 ms (low-rising: 0.88, 95% CI [0.81, 0.93]; flat: 0.63, 95% CI [0.55, 0.71]). For VOT at 0 ms, listeners gave predominantly /p/ responses and were influenced by F0 to a lesser degree than VOT at −20 ms (low: 0.17, 95% CI [0.11, 0.26]; flat: 0.07, 95% CI [0.04, 0.11]).

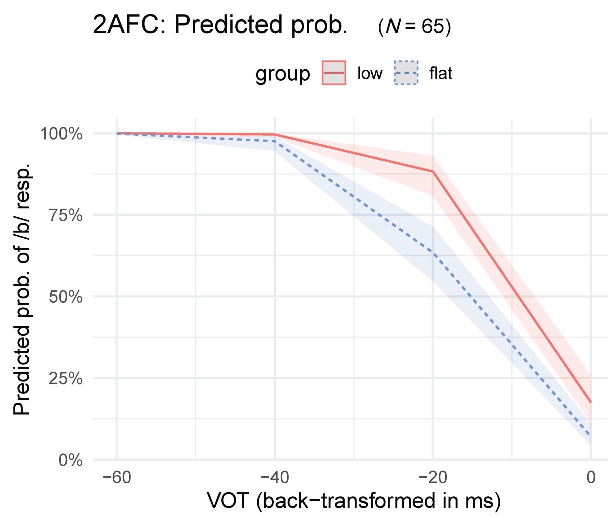

FIGURE 9. Prediction plot for /b/ response by VOT and F0 (model: *glmer*(*response* ~ *F*0 * *VOT* + (1 + *F*0 * *VOT*)|*subject*)). Shaded areas indicate 95% confidence intervals.

For some listeners, F0 has no effect, and some other listeners show no effect of VOT (see appendix C.2). It is unclear how to interpret these individual differences and how much should be explained by the remote collection setting, but the removal of these participants does not change the general outcomes of the model.

VAS responses. A beta regression GLMM model was fitted to the VAS scale responses [0, 100] rescaled to [0.01, 0.99], with orthogonal polynomial coded predictors (i.e. ordered categorical) VOT, F0, and their interaction. Random intercepts were included for participants, and by-participant random slopes for F0. The summary of the full model is given in appendix C.4.2. VOT and F0 were specified as ordered categorical rather than numerical predictors for two reasons. First, we sought to estimate the differences related to each F0 level at each VOT level. Second, the observed data in Fig. 8 suggest that the effect of F0 may be asymmetrical and/or unequal within different VOT ranges.

Post-hoc pairwise comparisons (Table 1) show the largest effects of F0 at −20 and 0 ms. We can observe that the effect of high (falling) and low (rising) F0 is not evenly distributed along the VOT continuum. Compared to flat F0, low F0 affects voicing identification only when VOT is at or above −20 ms, and its effect persists in the positive range. By contrast, the effect of high F0 is larger in the negative range and the largest at −20 ms. Taken together, the effect of F0 is especially observed in two situations: first, when VOT lies in the ambiguous region for voicing categorization, and second, when VOT and F0 are conflicting, that is, when the value of F0 is typically associated with the opposite category of the value of VOT (i.e. negative VOT with a high F0, or positive VOT with a low F0).

contrast	est	*SE*	*df*	*t*-ratio	*p*-value
VOT = −60					
100 - 130	−0.09	0.07	12,922	−1.352	0.3665
130 - 169	0.31	0.07	12,922	4.678	< 0.0001
VOT = −40					
100 - 130	−0.01	0.07	12,922	−0.091	0.9955
130 - 169	0.58	0.07	12,922	8.650	< 0.0001
VOT = −20					
100 - 130	0.34	0.07	12,922	5.069	< 0.0001
130 - 169	1.16	0.07	12,922	17.371	< 0.0001
VOT = 0					
100 - 130	0.59	0.07	12,922	8.673	< 0.0001
130 - 169	0.16	0.07	12,922	2.390	0.0444
VOT = 10					
100 - 130	0.49	0.07	12,922	7.146	< 0.0001
130 - 169	0.21	0.07	12,922	3.132	0.0049
VOT = 20					
100 - 130	0.49	0.07	12,922	7.102	< 0.0001
130 - 169	0.17	0.07	12,922	2.432	0.0399

Table 1. Estimated marginal means for the glmmTMB model predicting VAS response (Family: beta_family): *glmmTMB*(*response* ~ *as.ordered*(*VOT*) * *as.ordered*(*F*0) + (1 + *as.ordered*(*F*0))|*subject*). Results are given on the log-odds ratio (not the response) scale. *p*-value adjustment: Tukey method for comparing a family of three estimates. F0 = 100 for low-rising, 130 for flat, 169 for high-falling. The 100 - 169 pairs are not shown.

2AFC and VAS identification tasks: summary. The 2AFC and VAS identification tasks were conducted to see where on the VOT continuum onset F0 exerted the most influence on voicing category judgments. Onset F0 has the greatest influence around the category boundary, which is estimated to lie between −20 and 0 ms. In addition, onset F0 affects voicing categorization especially when VOT is ambiguous, or when VOT and F0 are conflicting. That is, the effect of onset F0 is not evenly distributed along the entire VOT continuum (RQ5). Except for the response curve with high-falling onset F0 showing somewhat continuous responses, listeners' responses in both tasks are quite categorical, despite previous findings showing that VAS tasks encourage gradient responses. This may be specific to the true-voicing contrast, or may be due to methodological limitations (McMurray 2022).

4. General discussion.

4.1. Trading between voicing and CF0: a result of categorization. The main goal of the current study was to investigate the relation between voicing and CF0 from the perspective of a frequently observed sound change, phonologization of co-intrinsic pitch (CF0). We tested whether the core assumptions of the CoPath model—namely inverse acoustic covariation and perceptual equivalence between coarticulatory source and effect—hold between voicing and CF0 in Metropolitan French, a true-voicing language where CF0 is not phonologized.

Our findings indicate that the correlation between VOT and onset F0 was absent or weakly positive within both voiced and voiceless categories (RQ1) as well as across categories (RQ2). In particular, in the production of /b/-onset words in isolation, the duration of prevoicing correlated weakly but positively with onset F0 lowering. In addition, we did not find lower onset F0 when devoicing occurred (RQ3), as might be expected if speakers were compensating for their failure to produce a canonical realization of the primary cue (§1.3). Taken together with previous findings from other languages, including Spanish (Dmitrieva et al. 2015), Tokyo Japanese (Gao & Arai 2019), and Central Swedish (Kirby & Tan 2023), our findings suggest that an inverse relation between coarticulatory source and effect does not apply to the production of voicing and CF0.

To answer RQ4, we conducted an AX discrimination experiment to test whether prevoicing and low onset F0 were perceptually equivalent when VOT was (i) unambiguously within the voiced category as well as (ii) near the voiced-voiceless boundary, that is, when the VOT cue was ambiguous in terms of voicing categorization. Our results revealed a critical effect of the ambiguity of VOT in this true-voicing language. When both stimuli clearly fell unambiguously within the voiced category, we found no difference in listeners' discrimination accuracy between the additive and canceling conditions. This is unexpected if perceptual equivalence would obtain between prevoicing and low onset F0 across a wide range of stimulus conditions. By contrast, when one stimulus had an ambiguous VOT (i.e. VOT was near the voiceless-voiced boundary), listeners' discrimination accuracy was higher for the additive than the canceling condition. That is, perceptually, the lowering of onset F0 could stand in for the absence of prevoicing, but not the temporal reduction of prevoicing. In sum, our AX discrimination results did not provide evidence for perceptual equivalence between prevoicing and low onset F0 within the voiced category. Instead, we observed a rather classic trading relation between these two cues only when the VOT cue was ambiguous.

To answer RQ5, we conducted two identification experiments to examine how onset F0 would affect listeners' voicing identification along the VOT continuum. In the 2AFC task, compared with a flat onset F0, a low-rising onset F0 lowered the voicing identification rate, especially near the voiceless-voiced boundary (i.e. when VOT was −20 or 0 ms). In the VAS task, the effect of onset F0 on voicing judgments was not evenly distributed along the VOT continuum. In the negative VOT range, high onset F0 had a stronger effect than low onset F0, whereas from −20 ms to the positive VOT range, low onset F0 had a stronger effect than high onset F0. In other words, listeners seemed to pay more attention to onset F0 when it conflicted with the category signaled by VOT than when it was congruent with that category.

Taken together, in production, we found no evidence for an inverse relation between voicing and CF0: if anything, we found these properties to be positively correlated. In perception, prevoicing appears to trade with low onset F0 only when VOT is ambiguous. Although onset F0 contributes to voicing judgments even when VOT is unambiguous (as previously shown by Whalen et al. 1993 for American English), its role is the greatest at the voiced-voiceless boundary. In addition, conflicting VOT-F0 combinations seem to have more impact on voicing judgments than nonconflicting combinations. These results suggest that, unlike vowel nasalization, the trading relation between voicing and CF0 is observed when a categorization decision is needed (Repp 1983), whereby the role of the secondary cue is especially important when the primary cue is ambiguous.

4.2. Spontaneous devoicing and perceptual adaptation. Recall that the CoPath model postulates that articulatory covariation and perceptual equivalence between coarticulatory source and effect serve as a path to sound change, whereby the coarticulatory effect becomes the dominant acoustic-perceptual property. In other words, the inverse relation between source and effect can be seen as a precursor of sound change (Beddor 2009:§2.3).

Our study shows an absence of a strictly inverse source-effect relation in the production and perception of voicing and CF0 in French, a true-voicing language. Does the absence of this precursor mean that the phonologization of CF0 would never take place in French? Perhaps. While the historical presence of canonical closure voicing appears to underlie many cases where CF0 has been phonologized, this does not mean that all languages with robust prevoicing must eventually transphonologize this property. However, our results may to some extent generalize to the case of voicing and CF0 in other languages. In the production of the voicing contrast in Afrikaans, younger speakers who have a much higher devoicing rate than older speakers nonetheless exhibit a similar temporal extent and magnitude of CF0 (Coetzee et al. 2018). While this does not mean that an inverse source-effect relation might not obtain in the general population, it suggests an absence of this relation at least across generations, in a language where there is an ongoing reweighting between voicing and CF0. In Tokyo Japanese, where obstruent devoicing is more frequent and CF0 is more pronounced than in French, the inverse source-effect relation is also absent (Gao & Arai 2019).

How, then, might the phonologization of CF0 come about in a true-voicing language like French? We propose that spontaneous devoicing acts as a natural production bias, resulting in a shift of perceptual attention to CF0. If listeners come to give CF0 more weight in perception and this subsequently comes to be reflected in their productions, the proportion of tokens with enhanced CF0 will increase over time, eventually leading to the phonologization of CF0. Below we elaborate on the details of this proposal.

As described in §1.3, the phonetic basis for the spontaneous devoicing of voiced obstruents is uncontroversial. Ohala (2011) elaborates two main options that languages have in dealing with the aerodynamic voicing constraint: 'let the AVC prevail' or 'circumvent the AVC'. The first option is simply to devoice phonetically voiced plosives. If the language's phonological inventory also contains voiceless obstruents, the diachronic consequence would be (potentially incomplete) merger or positional neutralization, such as syllable-final obstruent devoicing in German (Fourakis & Iverson 1984, Roettger et al. 2014) or word-initial fricative devoicing in some varieties of Dutch (Pinget et al. 2020). The second option is to attempt to preserve voicing by synergistic gestures, including but not limited to prenasalization, larynx lowering, or pharyngeal expansion (see §1.3). These multiple strategies are highly variable across languages and speakers. For example, in French and Spanish where voiced plosives are typically reported to be robustly prevoiced, speakers vary in the selection of one or a combination of above-mentioned articulatory maneuvers to circumvent the AVC in absolute initial position (Solé 2018). Depending on the particular strategy that prevails, this response may over time contribute to the creation of spirantized or prenasalized plosives, or implosives.

Here, we suggest a third possibility: 'try, but fail to circumvent the AVC'. Despite the speaker's best efforts to achieve closure voicing, glottal pulsing may still fail to occur if the requisite phonation pressure threshold is not met. However, because any F0 lowering would be the result of articulatory maneuvers undertaken to support voicing, any co-intrinsic effects on the F0 of the following vowels should persist, regardless of whether closure voicing actually occurs. Moreover, if the CF0 dichotomy is primarily driven by gestures that perturb F0 following voiceless plosives (which seems to be the case in French as well as in many other languages: Kirby & Ladd 2016), there is even less reason to predict that devoicing should disturb the CF0 dichotomy. Persistent devoicing of phonologically voiced plosives would thus result in a laryngeal contrast signaled solely by the postrelease differences in F0.

If, however, spontaneous devoicing does not simultaneously result in an enhancement of lowered onset F0 following devoiced plosives, as we have shown, why should this lead to the phonologization of CF0, rather than a merger with voiceless plosives? Along with previous findings, our study clearly shows that ambiguity in categorization created by devoiced plosives compels listeners to rely more on CF0 to identify the voicing category. We thus hypothesize that this shift of perceptual attention to CF0, if recurrent, may contribute to its phonologization. Previous work shows that listeners can be trained to shift their attention to a cue by means of short-term exposure to stimuli in which cue informativeness is increased (e.g. greater acoustic separation between categories: Francis et al. 2000, Holt & Lotto 2006, Idemaru & Holt 2014, Schertz & Clare 2020). Our recent work further suggests that when VOT becomes less informative, the RELATIVE informativeness of CF0 increases, creating favorable conditions for a short-term reallocation of perceptual attention to CF0 (Gao & Kirby 2023; see also Kim et al. 2020 for similar findings with vowel-related cues). However, it has also been shown that an individual's long-term cue weighting is stable and is dependent on the distributional regularities of the language environment (Idemaru et al. 2012). For this reason, we suspect that, for a persistent cue reweighting to become established, listeners would require long-term exposure to ambiguous items where the informativeness of prevoicing is reduced. Although this would not increase the informativeness of CF0 per se, upweighting of onset F0 (relative to VOT) would allow for a better mapping from listeners' phonological categorization to lexical and contextual feedback (Harmon et al. 2019). That is, persistent devoicing of phonologically voiced plosives would result in a laryngeal contrast signaled solely by the postrelease differences in F0, leading to a situation in which listeners would ONLY be able to rely on CF0 to make lexical decisions between minimal pairs involving devoiced and voiceless plosives. Indeed, our ongoing work (Gao & Kirby 2023) suggests an enhancing effect of lexical information on perceptual adaptation from voicing to CF0.

In sum, we propose that phonologization of CF0 begins with voicing failure contributing devoiced tokens to the pool of variation of phonologically voiced plosives (Ohala 1989), which may drive listeners' perceptual upweighting of CF0 for lexical disambiguation. Although we have not directly tested this idea in the present article, the findings presented here establish the empirical basis for this proposal: namely, that (a) devoicing does not impact the acoustic realization of CF0 in a true-voicing language, and (b) categorization ambiguity induced by devoicing increases the salience of CF0.

4.3. THEORETICAL PERSPECTIVES. Our proposal shares similarities and differences with the two models described in §1.1. Like Ohala (1981, 1989, 1993), our proposal highlights the crucial role of the listener, but we see the upweighting of CF0 as a reallocation of attentional resources rather than as an accidental misperception or misparsing, because we know of no evidence that CF0 is interpreted as a property of the vowel independently from its coarticulatory source. Moreover, we propose that this perceptual adaptation is guided by the need for perceptual distinctiveness to disambiguate members of a phonological or lexical contrast, a view to which Ohala (1981:186) does not subscribe. Like Beddor (2009) and Beddor et al. (2018), our proposal emphasizes the role of the speaker and the effect of phonetic environment on the reduction of coarticulatory source. But unlike in that model, we suggest that parity between production and perception is not mediated through a single shared gesture in the case of voicing and CF0. Indeed, our empirical findings differ from the vowel nasalization case investigated by Beddor (2009) in that a reduction of closure voicing (the coarticulatory source) does not result in an enhancement of CF0 (the, or at least a, coarticulatory effect).

As explained in §1.2, for vowel nasalization, both Beddor (2009) and Carignan et al. (2021) propose reduction of coarticulatory source as a precondition for phonologization of coarticulatory effect. We are proposing something similar here, namely, that spontaneous devoicing functions as a precursor for phonologization of CF0. The key difference is that, in the case of CF0, the coarticulatory source—the gestures that support or inhibit voicing—are not 'reduced' per se. As outlined above, failure to overcome the AVC does not entail the attenuation of these gestures, nor does it enhance the coarticulatory effect by itself. The enhancement of CF0—we hypothesize—is achieved through a perceptual adaptation strategy. While Carignan et al. (2021) also stress the role of shift of perceptual attention in the case of vowel nasalization, perceptual upweighting of the increased vowel nasality appears to operate in a gradual fashion, possibly mirroring the gradual increase of vowel nasality in production. In the case of voicing and CF0, our findings suggest that listeners attend to CF0 in a relatively categorical fashion, especially when the absence of closure voicing causes ambiguity in voicing categorization.

Setting aside these differences, the two factors we highlight here—production bias and perceptual adaptation—are shared in one form or another by all theories of sound change (for a classic review, see Garrett & Johnson 2013). In this way, the initiation of the phonologization of CF0 may be seen to follow a similar trajectory to that of vowel nasalization. At the same time, our findings are a reminder that different coarticulatory patterns may have very different source-effect and production-perception relationships. We also know that different coarticulatory patterns differ in terms of their regularity, typology, directionality, and so on. For example, while tonogenesis due to the phonologization of CF0 is diachronically well attested, the intrinsic F0 differences related to vowel height are rarely transphonologized (Arnold 2020, Hombert et al. 1979, Kingston 2011, Siddins & Harrington 2015, Ting et al. 2023), and 'tonoexodus' (i.e. tone loss) also seems to be relatively uncommon (Patin 2018, Ratliff 2015, Sims 2020). All of this means that we need to look in detail at the production and perception mechanisms for each coarticulatory pattern in order to model its behavior, direction, or rate in the course of sound change.

As a final remark, we stress that it is unlikely that the two factors we have highlighted here—devoicing and perceptual adaptation—can explain the phonologization of CF0 on their own. If both devoicing and perceptual adaptation are requirements for the subsequent enhancement of CF0 in production, the ongoing phonologization of CF0 in Afrikaans is unexpected, because older speakers exhibit a low rate of devoicing but nevertheless show substantial differences in CF0 (Coetzee et al. 2018). Other phonetic mechanisms are no doubt involved, such as larger effects of pitch raising than lowering in both production (Kang & Guion 2008, Kirby & Ladd 2016) and perception (Gao et al. 2019), the influence of place and manner of articulation (Hyslop 2022, Mazaudon 2012a, Peralta 2018), and the influence of phonation type (Gao et al. 2020, Henderson 1982, Pulleyblank 1978, Thurgood 2007). Furthermore, while we focus here on the initiation of sound change, it can be difficult to separate initiation from transmission and propagation (Hall-Lew et al. 2021). Structural and systemic factors (Brunelle & Kirby 2015, Brunelle & Pittayaporn 2012, Mazaudon 2012b, Michaud 2012), social factors (Yang et al. 2015), multilingualism, and contact (Evans 2001, Gao & Mazaudon 2022, Lau-Preechathammarach 2023, Pearce 2009, Steien & Yakpo 2020, Yeoung 2023) have been found or hypothesized to play a role in tonal developments and are likely to be operative at the earliest stages of tonogenetic sound changes.

5. Conclusion and outlook. Beddor (2009) proposes that an inverse relationship between coarticulatory source and effect in production and perception serves as a precursor for the phonologization of vowel nasalization. Inspired by her model, we looked for evidence of inverse acoustic covariation and perceptual equivalence between voicing and CF0 in Metropolitan French as a possible source of sound change. However, we found a different relationship between coarticulatory source and effect. Rather than an inverse correlation in production, we found at best a weakly positive correlation between the extent of prevoicing and onset F0. Rather than perceptual equivalence across the stimulus space, we found that the cue trading between prevoicing and onset F0 operates especially at the voiced-voiceless boundary.

We thus conclude that the phonologization of CF0 seen in 'tonogenetic' sound changes is driven not by an articulatorily based perceptual equivalence relation, as in the coarticulation of VN sequences. Instead, we suggest that spontaneous devoicing of onsets creates an ambiguity in voicing categorization that redirects listener attention to the secondary CF0 cue. Over time, through accumulated exposure to spontaneously devoiced voiced plosives, listeners may gradually come to upweight CF0 for lexical disambiguation. Subsequently, it is conceivable that such a perception pattern will lead to a substantial increase in the use of CF0 in production, establishing its phonologization. In ongoing work, we are testing the effect of accumulated exposure to ambiguous items coupled with lexical effects on listener-turned-speakers' reproduction of their perceptual pattern.

REFERENCES

Abramson, Arthur S., and Leigh Lisker. 1985. Relative power of cues: F0 shift versus voice timing. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, ed. by Victoria A. Fromkin, 25–33. New York: Academic Press.

Ahn, Suzy. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics* 71.451–67. DOI: 10.1016/j.wocn.2018.10.003.

Arnold, Laura. 2020. Highs and lows: A previously unattested tone split from vowel height in Metnyo Ambel. *Transactions of the Philological Society* 118.141–58. DOI: 10.1111/1467-968X.12177.

Bang, Hye-Young; Morgan Sonderegger; Yoonjung Kang; Meghan Clayards; and Tae-Jin Yoon. 2018. The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics* 66.120–44. DOI: 10.1016/j.wocn.2017.09.005.

Bates, Douglas; Martin Mächler; Ben Bolker; and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48. DOI: 10.18637/jss.v067.i01.

Beckman, Jill; Michael Jessen; and Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49.259–84. DOI: 10.1017/S0022226712000424.

Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85.785–821. DOI: 10.1353/lan.0.0165.

Beddor, Patrice Speeter. 2012. Perception grammars and sound change. *The initiation of sound change: Perception, production, and social factors*, ed. by Maria-Josep Solé and Daniel Recasens, 37–56. Amsterdam: John Benjamins.

Beddor, Patrice Speeter. 2023. Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change. *Journal of Phonetics* 98:101228. DOI: 10.1016/j.wocn.2023.101228.

Beddor, Patrice Speeter; Andries W. Coetzee; Will Styler; Kevin B. McGowan; and Julie E. Boland. 2018. The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language* 94.931–68. DOI: 10.1353/lan.2018.0051.

Bell-Berti, Fredericka. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. *The Journal of the Acoustical Society of America* 57.456–61. DOI: 10.1121/1.380468.

Benguerel, André-Pierre; Hajime Hirose; M. Sawashima; and T. Ushijima. 1978. Laryngeal control in French stop production: A fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica et Logopaedica* 30.175–98. DOI: 10.1159/000264124.

Best, Catherine T.; Barbara Morrongiello; and Rick Robson. 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics* 29.191–211. DOI: 10.3758/BF03207286.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486357.

Bradshaw, Mary Margaret. 1999. *A crosslinguistic study of consonant-tone interaction*. Columbus: The Ohio State University dissertation. Online: http://rave.ohiolink.edu/etdc/view?acc_num=osu1392896719.

Brooks, Mollie E.; Kasper Kristensen; Koen J. van Benthem; Arni Magnusson; Casper W. Berg; Anders Nielsen; Hans J. Skaug; Martin Mächler; and Benjamin M. Bolker. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9.378–400. DOI: 10.32614/RJ-2017-066.

Brunelle, Marc, and James Kirby. 2015. Re-assessing tonal diversity and geographical convergence in Mainland Southeast Asia. *Languages of Mainland Southeast Asia: The state of the art* (Pacific linguistics 649), ed. by N. J. Enfield and Bernard Comrie, 82–111. Berlin: De Gruyter Mouton. DOI: 10.1515/9781501501685-004.

Brunelle, Marc, and Pittayawat Pittayaporn. 2012. Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia. *Diachronica* 29.411–33. DOI: 10.1075/dia.29.4.01bru.

Carignan, Christopher; Stefano Coretta; Jens Frahm; Phil Hoole; Arun Joseph; Esther Kunay; and Dirk Voit. 2021. Planting the seed for sound change: Evidence from real-time MRI of velum kinematics in German. *Language* 97.333–64. DOI: 10.1353/lan.2021.0020.

Carignan, Christopher; Ryan Shosted; Chilin Shih; and Panying Rong. 2011. Compensatory articulation in American English nasalized vowels. *Journal of Phonetics* 39.668–82. DOI: 10.1016/j.wocn.2011.07.005.

Chen, Si; Yitian Hong; Bei Li; and Eunjin Chun. 2023. The f0 perturbation effects in focus marking: Evidence from Korean and Japanese. *PLOS ONE* 18:e0283139. DOI: 10.1371/journal.pone.0283139.

Chen, Yiya. 2011. How does phonology guide phonetics in segment-*f0* interaction? *Journal of Phonetics* 39.612–25. DOI: 10.1016/j.wocn.2011.04.001.

Chen, Yiya, and Laura J. Downing. 2011. All depressors are not alike: A comparison of Shanghai Chinese and Zulu. *Prosodic categories: Production, perception and comprehension*, ed. by Sónia Frota, Gorka Elordieta, and Pilar Prieto, 243–65. Dordrecht: Springer. DOI: 10.1007/978-94-007-0137-3_11.

Chi, Yujie; Kiyoshi Honda; and Jianguo Wei. 2019. Glottographic and aerodynamic analysis on consonant aspiration and onset F0 in Mandarin Chinese. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*, 6480–84. DOI: 10.1109/ICASSP.2019.8683882.

Choi, Jiyoun; Sahyang Kim; and Taehong Cho. 2020. An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLOS ONE* 15:e0240682. DOI: 10.1371/journal.pone.0240682.

Clayards, Meghan. 2018a. Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America* 144.EL172–EL177. DOI: 10.1121/1.5052025.

Clayards, Meghan. 2018b. Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica* 75.1–23. DOI: 10.1159/000448809.

Coetzee, Andries W.; Patrice Speeter Beddor; Kerby Shedden; Will Styler; and Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66.185–216. DOI: 10.1016/j.wocn.2017.09.009.

Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54.35–50. DOI: 10.1016/j.wocn.2015.09.003.

de Courtenay, Jan Baudouin. 1972 [1871]. *A Baudouin de Courtenay anthology: The beginnings of structural linguistics*. Ed. and trans. by Edward Stankiewicz. Bloomington: Indiana University Press. DOI: 10.2979/ABaudouindeCourtenay.

Di Cristo, Albert, and Daniel J. Hirst. 1986. Modelling French micromelody: Analysis and synthesis. *Phonetica* 43.11–30. DOI: 10.1159/000261758.

DiCanio, Christian, and Jared Sharp. 2020. Initial weakening in Mixtecan languages? *The Journal of the Acoustical Society of America* 148.2723. DOI: 10.1121/1.5147557.

Dmitrieva, Olga; Fernando Llanos; Amanda A. Shultz; and Alexander L. Francis. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49.77–95. DOI: 10.1016/j.wocn.2014.12.005.

Dockum, Rikker. 2019. *The tonal comparative method: Tai tone in historical perspective*. New Haven, CT: Yale University dissertation. Online: https://ling.yale.edu/sites/default/files/files/The_Tonal_Comparative_Method_.pdf.

Elliott, Emily M.; Raoul Bell; Simon Gorin; Nick Robinson; and John E. Marsh. 2022. Auditory distraction can be studied online! A direct comparison between in-person and online experimentation. *Journal of Cognitive Psychology* 34.307–24. DOI: 10.1080/20445911.2021.2021924.

Evans, Jonathan P. 2001. Contact-induced tonogenesis in Southern Qiang. *Language and Linguistics* 2.63–110. Online: https://www.ling.sinica.edu.tw/upload/researcher_manager_result/68419d0cb0e3f5adc38b8cfc973ae114.pdf.

Ferlus, Michel. 1979. Formation des registres et mutations consonantiques dans les langues mon-khmer. *Mon-Khmer Studies* 8.1–76. Online: http://sealang.net/sala/archives/pdf8/ferlus1979formation.pdf.

Fitch, Hollis L.; Terry Halwes; Donna M. Erickson; and Alvin M. Liberman. 1980. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics* 27.343–50. DOI: 10.3758/BF03206123.

Fourakis, Marios, and Gregory K. Iverson. 1984. On the ‘incomplete neutralization’ of German final obstruents. *Phonetica* 41.140–49. DOI: 10.1159/000261720.

Francis, Alexander L.; Kate Baldwin; and Howard C. Nusbaum. 2000. Effects of training on attention to acoustic cues. *Perception & Psychophysics* 62.1668–80. DOI: 10.3758/BF03212164.

Francis, Alexander L.; Valter Ciocca; Virginia Ka Man Wong; and Jess Ka Lam Chan. 2006. Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America* 120.2884–95. DOI: 10.1121/1.2346131.

Gao, Jiayin, and Takayuki Arai. 2019. Plosive (de-)voicing and *f*0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics* 77.1–33. DOI: 10.1016/j.wocn.2019.100932.

Gao, Jiayin; Pierre Hallé; and Christoph Draxler. 2020. Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception? *Language and Speech* 63.582–607. DOI: 10.1177/0023830919873080.

Gao, Jiayin, and James Kirby. 2023. Perceptual adaptation to altered cue informativeness: Distributional, auditory, and lexical factors. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, 161–65. Online: https://guarant.cz/icphs2023/179.pdf.

Gao, Jiayin, and Martine Mazaudon. 2022. Flexibility and evolution of cue weighting after a tonal split: An experimental field study on Tamang. *Linguistics Vanguard* 8.583–95. DOI: 10.1515/lingvan-2021-0085.

Gao, Jiayin; Jihyeon Yun; and Takayuki Arai. 2019. VOT-F0 coarticulation in Japanese: Production-biased or misparsing? *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)*, Melbourne, 210–14. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_259.pdf.

Gao, Jiayin; Jihyeon Yun; and Takayuki Arai. 2021. Korean laryngeal contrast revisited: An electroglottographic study on denasalized and oral stops. *Laboratory Phonology* 12(1):7. DOI: 10.5334/labphon.277.

Garner, Wendell R. 1974. *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum.

Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. In Yu 2013, 51–97. DOI: 10.1093/acprof:oso/9780199573745.003.0003.

Gay, Thomas; Björn Lindblom; and James Lubker. 1981. Production of bite-block vowels: Acoustic equivalence by selective compensation. *Journal of the Acoustical Society of America* 69.802–10. DOI: 10.1121/1.385591.

Gehrmann, Ryan. 2022. *Desegmentalization: Towards a common framework for the modeling of tonogenesis and registrogenesis in Mainland Southeast Asia with case studies from Austroasiatic*. Edinburgh: University of Edinburgh dissertation. DOI: 10.7488/era/2364.

Goldstein, Louis. 1990. On articulatory binding: Comments on Kingston's paper. *Papers in laboratory phonology 1: Between the grammar and physics of speech*, ed. by John Kingston and Mary E. Beckman, 445–50. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627736.025.

Gordon, Matthew. 2016. Consonant-tone interactions: A phonetic study of four indigenous languages of the Americas. *The phonetics and phonology of laryngeal features in native American languages*, ed. by Heriberto Avelino, Matt Coler, and Leo Wetzels, 129–56. Leiden: Brill. DOI: 10.1163/9789004303218_006.

Guo, Yuting, and Harim Kwon. 2022. Production and perception of Mandarin laryngeal contrast: The role of post-plosive F0. *Frontiers in Communication* 7:896013. DOI: 10.3389/fcomm.2022.896013.

Hagège, Claude, and André Haudricourt. 1978. *La phonologie panchronique : Comment les sons changent dans les langues*. Paris: Presses Universitaires de France.

Haggard, Mark; Stephen Ambler; and Mo Callow. 1970. Pitch as a voicing cue. *The Journal of the Acoustical Society of America* 47.613–17. DOI: 10.1121/1.1911936.

Hall-Lew, Lauren; Patrick Honeybone; and James Kirby. 2021. Individuals, communities, and sound change: An introduction. *Glossa: a journal of general linguistics* 6(1):67. DOI: 10.5334/gjgl.1630.

Halle, Morris, and Kenneth N. Stevens. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Research Report* 101.198–213.

Hanson, Helen M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America* 125.425–41. DOI: 10.1121/1.3021306.

Harmon, Zara; Kaori Idemaru; and Vsevolod Kapatsinski. 2019. Learning mechanisms in cue reweighting. *Cognition* 189.76–88. DOI: 10.1016/j.cognition.2019.03.011.

Harrington, Jonathan; Felicitas Kleber; Ulrich Reubold; Florian Schiel; and Mary Stevens. 2019. The phonetic basis of the origin and spread of sound change. *The Routledge handbook of phonetics*, ed. by William F. Katz and Peter F. Assmann, 401–26. New York: Routledge.

Haudricourt, André G. 1961. Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient. *Bulletin de la Société de Linguistique de Paris* 56.163–80.

Henderson, Eugénie J. A. 1982. Tonogenesis: Some recent speculations on the development of tone. *Transactions of the Philological Society* 80.1–24. DOI: 10.1111/j.1467-968X.1982.tb01192.x.

Holt, Lori L., and Andrew J. Lotto. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America* 119.3059–71. DOI: 10.1121/1.2188377.

Hombert, Jean-Marie. 1976. The effect of aspiration on the fundamental frequency of the following vowel. *Berkeley Linguistics Society* 2.212–19. Online: https://escholarship.org/uc/item/0w7198mj.

Hombert, Jean-Marie; John J. Ohala; and William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55.37–58. DOI: 10.2307/412518.

Honda, Kiyoshi; Hiroyuki Hirai; Shinobu Masaki; and Yasuhiro Shimada. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42.401–11. DOI: 10.1177/00238309990420040301.

Honeybone, Patrick. 2005. Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. *The internal organization of phonological segments*, ed. by Marc van Oostendorp and Jeroen van de Weijer, 319–54. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110890402.317.

Hoole, Philip. 2006. *Experimental studies of laryngeal articulation*. Munich: University of Munich habilitation thesis. Online: https://www.phonetik.uni-muenchen.de/~hoole/pdf/habilemg_chap_all.pdf.

Hoole, Philip, and Kiyoshi Honda. 2011. Automaticity vs. feature-enhancement in the control of segmental F0. *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories*, ed. by G. Nick Clements and Rachid Ridouane, 131–72. Amsterdam: John Benjamins.

House, Arthur S., and Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustic characteristics of vowels. *The Journal of the Acoustical Society of America* 25.105–13. DOI: 10.1121/1.1906982.

Howe, Penelope Jane. 2017. *Tonogenesis in Central dialects of Malagasy: Acoustic and perceptual evidence with implications for synchronic mechanisms of sound change*. Houston: Rice University dissertation. Online: https://hdl.handle.net/1911/96031.

Hutin, Mathilde; Yaru Wu; Adèle Jatteau; Ioana Vasilescu; Lori Lamel; and Martine Adda-Decker. 2021. Synchronic fortition in five Romance languages? A large corpus-based study of word-initial devoicing. *Proceedings of Interspeech 2021*, 996–1000. DOI: 10.21437/Interspeech.2021-939.

Hyman, Larry M. 1976. Phonologization. *Linguistic studies offered to Joseph Greenberg*, ed. by Alphonse G. Juilland, 407–41. Saratoga, CA: Anma Libri.

Hyman, Larry M. 2013. Enlarging the scope of phonologization. In Yu 2013, 3–28. DOI: 10.1093/acprof:oso/9780199573745.003.0001.

Hyslop, Gwendolyn. 2022. Toward a typology of tonogenesis: Revising the model. *Australian Journal of Linguistics* 42.275–99. DOI: 10.1080/07268602.2022.2157675.

Idemaru, Kaori, and Lori L. Holt. 2011. Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance* 37.1939–56. DOI: 10.1037/a0025641.

Idemaru, Kaori, and Lori L. Holt. 2014. Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 40.1009–21. DOI: 10.1037/a0035269.

Idemaru, Kaori; Lori L. Holt; and Howard Seltman. 2012. Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America* 132.3950–64. DOI: 10.1121/1.4765076.

Iverson, Gregory K., and Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12.369–96. DOI: 10.1017/S0952675700002566.

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody*. Columbus: The Ohio State University dissertation. Online: http://rave.ohiolink.edu/etdc/view?acc_num=osu1220465077.

Kang, Kyoung-Ho, and Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America* 124.3909–17. DOI: 10.1121/1.2988292.

Kang, Yoonjung, and Manami Hirayama. 2023. Age- and gender-based variation in the perception of voicing contrast in Tokyo Japanese. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, 3026–30. Online: https://guarant.cz/icphs2023/509.pdf.

Kapnoula, Efthymia C.; Matthew B. Winn; Eun Jong Kong; Jan Edwards; and Bob McMurray. 2017. Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance* 43.1594–1611. DOI: 10.1037/xhp0000410.

Keating, Patricia A.; Wendy Linker; and Marie Huffman. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11.277–90. DOI: 10.1016/S0095-4470(19)30827-7.

Keyser, Samuel Jay, and Kenneth Noble Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82.33–63. DOI: 10.1353/lan.2006.0051.

Kim, Donghyun; Meghan Clayards; and Eun Jong Kong. 2020. Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics* 81:100984. DOI: 10.1016/j.wocn.2020.100984.

Kim, Hyunsoon; Shinji Maeda; Kiyoshi Honda; and Lise Crevier-Buchman. 2018. The mechanism and representation of Korean three-way phonation contrast: External photoglottography, intra-oral air pressure, airflow, and acoustic data. *Phonetica* 75.57–84. DOI: 10.1159/000479589.

Kingston, John. 1990. Articulatory binding. *Papers in laboratory phonology 1: Between the grammar and physics of speech*, ed. by John Kingston and Mary E. Beckman, 406–34. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627736.023.

Kingston, John. 2007. Segmental influences on F0: Automatic or controlled? *Tones and tunes, vol. 2: Experimental studies in word and sentence prosody*, ed. by Carlos Gussenhoven and Tomas Riad, 171–201. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110207576.2.171.

Kingston, John. 2011. Tonogenesis. *The Blackwell companion to phonology, vol. 4: Phonological interfaces*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, and Keren Rice, 2304–33. Oxford: Blackwell. DOI: 10.1002/9781444335262.wbctp0097.

Kingston, John, and Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70.419–54. DOI: 10.2307/416481.

Kingston, John; Randy L. Diehl; Cecilia J. Kirk; and Wendy A. Castleman. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics* 36.28–54. DOI: 10.1016/j.wocn.2007.02.001.

Kirby, James P. 2013. The role of probabilistic enhancement in phonologization. In Yu 2013, 228–46. DOI: 10.1093/acprof:oso/9780199573745.003.0011.

Kirby, James P. 2018. Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics* 71.326–54. DOI: 10.1016/j.wocn.2018.09.009.

Kirby, James P., and D. Robert Ladd. 2015. Stop voicing and F0 perturbations: Evidence from French and Italian. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0740.pdf.

Kirby, James P., and D. Robert Ladd. 2016. Effects of obstruent voicing on vowel F0: Evidence from 'true voicing' languages. *The Journal of the Acoustic Society of America* 140.2400–2411. DOI: 10.1121/1.4962445.

Kirby, James P.; D. Robert Ladd; Jiayin Gao; and Zuzana Elliott. 2020. Elicitation context does not drive F0 lowering following voiced stops: Evidence from French and Italian. *The Journal of the Acoustical Society of America* 148.EL147–EL152. DOI: 10.1121/10.0001698.

Kirby, James, and Maryann Tan. 2023. Analyzing variability in closure voicing and co-intrinsic F0 in Central Standard Swedish. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, 2268–72. Online: https://guarant.cz/icphs2023/356.pdf.

Kohler, Klaus J. 1982. F0 in the perception of lenis and fortis plosives. *Phonetica* 39.199–218. DOI: 10.1159/000261663.

Kong, Eun Jong, and Jan Edwards. 2016. Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics* 59.40–57. DOI: 10.1016/j.wocn.2016.08.006.

Krug, Paul Konstantin; Branislav Gerazov; Daniel R. van Niekerk; Anqi Xu; Yi Xu; and Peter Birkholz. 2021. Modelling microprosodic effects can lead to an audible improvement in articulatory synthesis. *The Journal of the Acoustical Society of America* 150.1209–17. DOI: 10.1121/10.0005876.

Kuang, Jianjing, and Aletheia Cui. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71.194–214. DOI: 10.1016/j.wocn.2018.09.002.

Kühnert, Barbara; C. Ledl; Philip Hoole; and H. G. Tillmann. 1991. Tongue-jaw interactions in lingual consonants. *PERILUS (Phonetic Experimental Research at the Institute of Linguistics University of Stockholm)* 14.21–25. Online: https://www.phonetik.uni-muenchen.de/~hoole/pdf/kuehnertetal_perilus_14_1991.pdf.

Ladefoged, Peter. 1973. The features of the larynx. *Journal of Phonetics* 1.73–83. DOI: 10.1016/S0095-4470(19)31376-2.

Lau-Preechathammarach, Raksit Tyler. 2023. The interaction between language usage and acoustic correlates of the Kuy register distinction. *Laboratory Phonology* 14(1). DOI: 10.16995/labphon.6531.

Lee, Seunghun Julio. 2008. *Consonant-tone interaction in optimality theory*. New Brunswick, NJ: Rutgers University dissertation. DOI: 10.7282/T32F7NS1.

Lenth, Russell V. 2021. emmeans: Estimated marginal means, aka least-squares means. R package. Online: https://cran.r-project.org/package=emmeans.

Li, Fang Kuei. 1980. Laryngeal features and tone development. *Bulletin of the Institute of History and Philology Academia Sinica* 51.1–13. Online: http://www2.ihp.sinica.edu.tw/file/3595FkIBSwe.pdf.

Liljencrants, Johan, and Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48.839–62. DOI: 10.2307/411991.

Lisker, Leigh. 1975. Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America* 57.1547–51. DOI: 10.1121/1.380602.

Lisker, Leigh. 1986. ‘Voicing’ in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29.3–11. DOI: 10.1177/002383098602900102.

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20.384–422. DOI: 10.1080/00437956.1964.11659830.

Lisker, Leigh; Alvin M. Liberman; Donna M. Erickson; David Dechovitz; and Roland Mandler. 1977. On pushing the voice-onset-time (VOT) boundary about. *Language and Speech* 20.209–16. DOI: 10.1177/002383097702000303.

Llanos, Fernando; Olga Dmitrieva; Amanda Shultz; and Alexander L. Francis. 2013. Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America* 134.2213–24. DOI: 10.1121/1.4817845.

Löfqvist, Anders; Thomas Baer; Nancy S. McGarr; and Robin Seider Story. 1989. The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America* 85.1314–21. DOI: 10.1121/1.397462.

Luo, Qian. 2018. *Consonantal effects on F0 in tonal languages*. East Lansing: Michigan State University dissertation. DOI: 10.25335/M5N58CQ8J.

Macmillan, Neil A., and C. Douglas Creelman. 2005. *Detection theory: A user’s guide*. 2nd edn. London: Routledge.

Maeda, Shinji. 1990. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. *Speech production and speech modelling*, ed. by William J. Hardcastle and Alain Marchal, 131–49. Dordrecht: Springer. DOI: 10.1007/978-94-009-2037-8_6.

Mann, Virginia A. 1980. Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics* 28.407–12. DOI: 10.3758/BF03204884.

Mann, Virginia A., and Bruno H. Repp. 1981. Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America* 69.548–58. DOI: 10.1121/1.385483.

Maspero, Henri. 1912. Études sur la phonétique historique de la langue annamite: Les initiales [Studies in Annamese historical phonetics: Initial consonants]. *Bulletin de l’École française d’Extrême-Orient* 12.1–127. Online: https://www.jstor.org/stable/43729467.

Massaro, Dominic W., and Michael M. Cohen. 1983. Categorical or continuous speech perception: A new test. *Speech Communication* 2.15–35. DOI: 10.1016/0167-6393(83)90061-4.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. *Consonant types and tone*, ed. by Larry Hyman, 71–95. Los Angeles: Linguistics Program of USC.

Mazaudon, Martine. 2012a. The influence of tone and affrication on manner: Some irregular manner correspondences in the Tamang group. *Linguistics of the Tibeto-Burman Area* 35.97–112. Online: http://sealang.net/archives/ltba/pdf/LTBA-35.2.97.pdf.

Mazaudon, Martine. 2012b. Paths to tone in the Tamang branch of Tibeto-Burman (Nepal). *The dialect laboratory: Dialects as a testing ground for theories of language change*, ed. by Gunther De Vogelaer and Guido Seiler, 139–75. Amsterdam: John Benjamins.

Mazaudon, Martine, and Alexis Michaud. 2008. Tonal contrasts and initial consonants: A case study of Tamang, a 'missing link' in tonogenesis. *Phonetica* 65.231–56. DOI: 10.1159/000192794.

McMurray, Bob. 2022. The myth of categorical perception. *The Journal of the Acoustical Society of America* 152.3819–42. DOI: 10.1121/10.0016614.

Michaud, Alexis. 2012. Monosyllabicization: Patterns of evolution in Asian languages. *Monosyllables: From phonology to typology*, ed. by Thomas Stolz, Nicole Nau, and Cornelia Stroh, 115–30. Berlin: Akademie. DOI: 10.1524/9783050060354.115.

Michaud, Alexis, and Bonny Sands. 2020. Tonogenesis. *The Oxford research encyclopedia of linguistics*, ed. by Mark Aronoff. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.748.

Moisik, Scott R.; Ewa Czaykowska-Higgins; and John H. Esling. 2021. Phonological potentials and the lower vocal tract. *Journal of the International Phonetic Association* 51.1–35. DOI: 10.1017/S0025100318000403.

Ohala, John J. 1981. The listener as a source of sound change. *Chicago Linguistic Society (Parasession on language and behavior)* 17(2).178–203.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. *The production of speech*, ed. by Peter F. MacNeilage, 189–216. Dordrecht: Springer. DOI: 10.1007/978-1-4613-8202-7_9.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, ed. by Leiv E. Breivik and Ernst Håkon Jahr, 173–98. Berlin: De Gruyter. DOI: 10.1515/9783110853063.173.

Ohala, John J. 1993. The phonetics of sound change. *Historical linguistics: Problems and perspectives*, ed. by Charles Jones, 237–78. London: Longman.

Ohala, John J. 2011. Accommodation to the aerodynamic voicing constraint and its phonological relevance. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 64–67. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/SpecialSession/Session1/Ohala/Ohala.pdf.

Ohde, Ralph N. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America* 75.224–30. DOI: 10.1121/1.390399.

Pape, Daniel, and Luis M. T. Jesus. 2015. Stop and fricative devoicing in European Portuguese, Italian and German. *Language and Speech* 58.224–46. DOI: 10.1177/0023830914530604.

Pardo, Jennifer S., and Carol A. Fowler. 1997. Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception & Psychophysics* 59.1141–52. DOI: 10.3758/BF03205527.

Passy, Paul. 1890. *Étude sur les changements phonétiques*. Paris: Librarie Firmin-Didot.

Patin, Cédric. 2018. From tones to accents. *Proceedings of the 6th International Symposium on Tonal Aspects of Language (TAL 2018)*, 17–21. DOI: 10.21437/TAL.2018-4.

Pearce, Mary. 2009. Kera tone and voicing interaction. *Lingua* 119.846–64. DOI: 10.1016/j.lingua.2007.10.023.

Peralta, William. 2018. Tonogenesis: The perception of tone and the role of place of articulation in Kurtöp. *Proceedings of the 6th International Symposium on Tonal Aspects of Languages (TAL 2018)*, 83–87. DOI: 10.21437/TAL.2018-17.

Perkell, Joseph S.; Melanie L. Matthies; Mario A. Svirsky; and Michael I. Jordan. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study. *The Journal of the Acoustical Society of America* 93.2948–61. DOI: 10.1121/1.405814.

Perrier, Pascal, and Susanne Fuchs. 2015. Motor equivalence in speech production. *The handbook of speech production*, ed. by Melissa A. Redford, 225–47. London: Wiley-Blackwell. DOI: 10.1002/9781118584156.ch11.

Pinget, Anne-France; René Kager; and Hans Van de Velde. 2020. Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech* 63.660–85. DOI: 10.1177/0023830919880206.

Pittayaporn, Pittayawat. 2009. *The phonology of Proto-Tai*. Ithaca, NY: Cornell University dissertation. Online: https://hdl.handle.net/1813/13855.

Premsrirat, Suwilai. 2001. Tonogenesis in Khmu dialects of Southeast Asia. *Papers from the ninth annual meeting of the Southeast Asian Linguistics Society*, 121–34. Online: http://sealang.net/sala/archives/pdf8/premsrirat2001tonogenesis.pdf.

Pulleyblank, Edwin G. 1978. The nature of the Middle Chinese tones and their development to Early Mandarin. *Journal of Chinese Linguistics* 6.173–203. Online: https://www.jstor.org/stable/23752830.

R Core Team. 2021. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: https://www.r-project.org/.

Ratliff, Martha. 2015. Tonoexodus, tonogenesis, and tone change. *The Oxford handbook of historical phonology*, ed. by Patrick Honeybone and Joseph Salmons, 245–61. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199232819.013.021.

Reinholt Petersen, Niels. 1983. The effect of consonant type on fundamental frequency and larynx height in Danish. *Annual Report of the Institute of Phonetics University of Copenhagen* 17.55–86. DOI: 10.7146/aripuc.v17i.131784.

Repp, Bruno H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92.81–110. DOI: 10.1037/0033-2909.92.1.81.

Repp, Bruno H. 1983. Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication* 2.341–61. DOI: 10.1016/0167-6393(83)90050-X.

Roettger, Timo B.; Bodo Winter; Sven Grawunder; James Kirby; and Martine Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43.11–25. DOI: 10.1016/j.wocn.2014.01.002.

Sanker, Chelsea. 2023. How do headphone checks impact perception data? *Laboratory Phonology* 14(1). DOI: 10.16995/labphon.8778.

Schertz, Jessamyn. 2014. Klatt synthesize VOT-F0 series. Online: http://individual.utoronto.ca/jschertz/scripts/klatt_synthesize_vot_f0_series.praat.

Schertz, Jessamyn, and Emily J. Clare. 2020. Phonetic cue weighting in perception and production. *WIREs Cognitive Science* 11:e1521. DOI: 10.1002/wcs.1521.

Serniclaes, Willy. 1987. *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Brussels: Université Libre de Bruxelles dissertation. Online: http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/213529.

Shi, Menghui; Yiya Chen; and Maarten Mous. 2020. Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese. *The Journal of the Acoustical Society of America* 147.2901–16. DOI: 10.1121/10.0001000.

Shimizu, Katsumasa. 1990. *Cross-language study of voicing contrasts of stop consonants in Asian languages*. Edinburgh: University of Edinburgh dissertation. Online: http://hdl.handle.net/1842/20185.

Shultz, Amanda A.; Alexander L. Francis; and Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132.EL95–EL101. DOI: 10.1121/1.4736711.

Siddins, Jessica, and Jonathan Harrington. 2015. Does vowel intrinsic f0 affect lexical tone? *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0678.pdf.

Silverman, Kim. 1986. F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43.76–91. DOI: 10.1159/000261762.

Simon, Claude, and Adrian J. Fourcin. 1978. Cross-language study of speech-pattern learning. *The Journal of the Acoustical Society of America* 63.925–35. DOI: 10.1121/1.381772.

Sims, Nathaniel A. 2020. Reconsidering the diachrony of tone in Rma. *Journal of the Southeast Asian Linguistics Society* 13.53–85. Online: http://hdl.handle.net/10524/52460.

Solé, Maria-Josep. 2018. Articulatory adjustments in initial voiced stops in Spanish, French and English. *Journal of Phonetics* 66.217–41. DOI: 10.1016/j.wocn.2017.10.002.

Soli, Sigfrid D. 1983. The role of spectral cues in discrimination of voice onset time differences. *The Journal of the Acoustical Society of America* 73.2150–65. DOI: 10.1121/1.389539.

Stanislaw, Harold, and Natasha Todorov. 1999. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* 31.137–49. DOI: 10.3758/BF03207704.

Steien, Guri Bordal, and Kofi Yakpo. 2020. Romancing with tone: On the outcomes of prosodic contact. *Language* 96.1–41. DOI: 10.1353/lan.2020.0000.

Stevens, Kenneth N., and Dennis H. Klatt. 1974. Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America* 55.653–59. DOI: 10.1121/1.1914578.

Stoet, Gijsbert. 2010. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42.1096–1104. DOI: 10.3758/BRM.42.4.1096.

Stoet, Gijsbert. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44.24–31. DOI: 10.1177/0098628316677643.

Svantesson, Jan-Olof, and David House. 2006. Tone production, tone perception and Kammu tonogenesis. *Phonology* 23.309–33. DOI: 10.1017/S0952675706000923.

Takada, Mieko; Eun Jong Kong; Kiyoko Yoneyama; and Mary E. Beckman. 2015. Loss of prevoicing in modern Japanese /g, d, b/. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0873.pdf.

Tang, Katrina Elizabeth. 2008. *The phonology and phonetics of consonant-tone interaction*. Los Angeles: University of California, Los Angeles dissertation. Online: https://linguistics.ucla.edu/general/dissertations/TangPhonologyPhoneticsConsonantToneInteraction_small.pdf.

Thurgood, Graham. 2007. Tonogenesis revisited: Revising the model and the analysis. *Studies in Tai and Southeast Asian linguistics*, ed. by Jimmy G. Harris, Somsonge Burusphat, and James E. Harris, 263–91. Bangkok: Ekphimthai.

Ting, Connie; Meghan Clayards; Morgan Sonderegger; and Michael McAuliffe. 2023. The cross-linguistic distribution of vowel and consonant intrinsic f0 effects. Montreal: McGill University, ms. Online: https://psyarxiv.com/64nhs/.

Titze, Ingo R. 2000. *Principles of voice production*. 2nd printing edn. Iowa City: National Center for Voice and Speech.

Weenink, David. 2009. The KlattGrid speech synthesizer. *Proceedings of Interspeech 2009*, 2059–62. DOI: 10.21437/Interspeech.2009-591.

Werker, Janet F., and John S. Logan. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37.35–44. DOI: 10.3758/BF03207136.

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America* 73.1322–36. DOI: 10.1121/1.389236.

Whalen, Douglas H.; Arthur S. Abramson; Leigh Lisker; and Maria Mody. 1990. Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47.36–49. DOI: 10.1159/000261851.

Whalen, Douglas H.; Arthur S. Abramson; Leigh Lisker; and Maria Mody. 1993. *F*0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America* 93.2152–59. DOI: 10.1121/1.406678.

Xu, Ching X., and Yi Xu. 2003. Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* 33.165–81. DOI: 10.1017/S0025100303001270.

Xu, Yi, and Anqi Xu. 2021. Consonantal F_0 perturbation in American English involves multiple mechanisms. *The Journal of Acoustical Society of America* 149.2877–95. DOI: 10.1121/10.0004239.

Yang, Cathryn; James N. Stanford; and Zhengyu Yang. 2015. A sociotonetic study of Lalo tone split in progress. *Asia-Pacific Language Variation* 1.52–77. DOI: 10.1075/aplv.1.1.03yan.

Yeoung, Ping Hei. 2023. Contact-induced tonogenesis in Hong Kong English. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, 1945–49. Online: https://guarant.cz/icphs2023/440.pdf.

Yu, Alan C. L. (ed.) 2013. *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199573745.001.0001.

Yu, Alan C. L. 2021. Toward an individual-difference perspective on phonologization. *Glossa: a journal of general linguistics* 6(1):14. DOI: 10.5334/gjgl.661.

Zhang, Yubin, and Louis Goldstein. 2023. Stop voicing and devoicing as articulatory tasks: A cross-linguistic RT-MRI study. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, 1011–15. Online: https://guarant.cz/icphs2023/416.pdf.

[jiayin.gao@cnrs.fr]
[jkirby@phonetik.uni-muenchen.de]

[Received 27 April 2022;
revision invited 28 August 2022;
revision received 10 May 2023;
accepted pending revisions 20 August 2023;
revision received 16 September 2023;
accepted 21 September 2023]