Toward an individual-difference perspective on phonologization

Alan C. L. Yu

University of Chicago
1115 E. 58th Street, Chicago, IL 60637
aclyu@uchicago.edu

Word count: 10293

Abstract Phonologization is often understood to be a process along the pathway of sound change where low-level physiological or perceptual variation that gives rise to sound patterns is explicitly encoded in the grammar. The nature of this transition and how it comes about remain largely unclear. The broad theoretical goal of this paper is to argue that phonologization should be defined at the level of the individual; it takes place whenever an individual acquires a so-called "phonetic precursor" as an intended, and controlled, pattern in the language, even if the pattern exhibits gradient properties. To illustrate this perspective, a case study on the nature of individual variation in vowel duration as a function of vowel height in Cantonese is presented, offering evidence that some Cantonese speakers have phonologized the height-dependent vowel duration effect and discussing the implications of this result for understanding intrinsic vowel duration as a phonetic precursor to sound change. The article closes with a discussion regarding the current proposal in relation to past models of phonologization and offers suggestions for future directions in phonologization research.

Keywords: individual difference, intrinsic vowel duration, phonologization, sound change

1   Introduction

Phonologization is often understood to be a process along the pathway of sound change where low-level physiological or perceptual variation that gives rise to sound patterns is explicitly encoded in the grammar. The notion of phonologization, while often seen as a central component of any theory of sound change, is also one of the most contested. The interpretation of the term can differ widely depending on one's assumptions about the division (or the lack thereof) between phonetics and phonology, and the conception of sound change in general. Nonetheless, all approaches to phonologization are inspired by the strong parallels between synchronic phonetic variations and diachronic sound changes. The nature of this parallelism and how it is related to phonologization remain a matter of debate, however. Specifically, what is the nature of this transition and how it comes about remain largely unclear.

   Cue reweighting approaches to sound change, for example, assume that intrinsic phonetic variation must be "exaggerated" to become phonologized. In his seminal study on phonologization, Hyman (1976) explained the emergence of allophonic pitch variation as a result of the phonologization of consonantal perturbation of pitch on the neighboring vowel, as a result of the physiological contingencies for producing obstruent voicing. In particular, the pitch perturbation is said to be exaggerated to such an extent that the

pitch variation cannot be attributed entirely to the physiological properties of the preceding consonant's voicing and must be included as part of the phonology of the language (e.g., *pa > pá and *ba > bà; á is a vowel with high tone, while à is a vowel with low tone). This notion of the "exaggeration" of an intrinsic phonetic precursor has remained under-articulated, however (cf. Hamann 2009; Ramsammy 2015; 2018; see also discussion below). In particular, if the phonetic preconditions for change are always present, as they are attributed to universal phonetic principles, what accounts for the sudden exaggeration of the phonetic precursor?

Other approaches emphasize the effects perceptual parsing has on sound change (Blevins 2004; Beddor 2009; Ohala 1981; 1983; 1989; 1992; 1993; 1995). That is, the ambiguous nature of the speech signal, which presumably stems from articulatory, acoustic, auditory, and perceptual constraints inherent to the vocal tract along with the auditory and perceptual apparatus, is seen as the main culprit behind the emergence of new sound patterns. For example, in the case of the development of contrastive nasal vowels from vowel + nasal sequences (VN > Ṽ), a sound change frequently attested in the world's languages, listeners must discern the provenance of nasalization present on the vowel. Theorists within this tradition often assume that listeners regularly factor out (i.e. perceptually compensate for) the presence of vowel nasalization as an artifact of anticipatory coarticulation. However, occasionally, the listeners might interpret the presence of nasalization as a part of the intended speech signal, which can lead to a mini-sound change. The listeners might also analyze nasal "prosodically" and assume a perceptual equivalence between the nasality of the vowel and the nasality of the nasal consonant (Beddor 2009). From this perspective, phonologization "involves a stage of coarticulatory variation in which the duration of the coarticulatory source and the temporal extent of its influence on nearby segments are inversely related (Beddor 2009: 813)."[1] The misperception/reanalysis-driven approaches to sound change, to the extent it is discussed explicitly, often see the stabilization of a phonetic precursor into a sound pattern as the result of the accumulation of experience. Ohala, for example, sees the emergence of novel variants as lexically gradual. That is, phonologization happens as a result of misperception or reanalysis one word at a time. It is through the accumulation of such misperceived or reanalyzed tokens that systemic shifts are realized (see Bermúdez-Otero 2015 for an explicit discussion regarding a token-accumulation account for the emergence of vowel backing). As systematic changes are unlikely to arise from haphazard perceptual mistakes, children and L2 learners are seen as the drivers behind listener misperception-derived sound changes (Ohala 1993). They are presumed to have strong preconceptions about the language in the first place. However, as noted by various scholars (e.g., Yu 2007; Stevens & Harrington 2014), novel variants arisen from children and L2 learners are unlikely to lead to sound change at the community level because native

---

[1] Beddor (2009) couched the so-called "coarticulatory path to sound change" in terms of speaker-listener assigning different perceptual weighting to acoustic-auditory properties that map to that representation. However, unlike the cue weighting approach to sound change reviewed above, the perceptual equivalence pathway to phonologization is not simply a matter of up- or down-weighting a cue's role in perception, as in the case of the development of allotony out of f0-perturbation due to stop voicing (Kirby 2010), but about how the temporal organization of cues is analyzed and understood by the speaker-listener. That is, at the stage when nasality is phonologized on the vowel in a VN sequence, the temporal span of nasality is no longer understood to be associated solely with the nasal stop, but is reanalyzed as encompassing a domain that encompasses two segments; nasality is analyzed prosodically or as an "autosegment" in Autosegmental Phonology (Goldsmith 1990).

listeners would likely identify "innovations" from such inexperienced speakers as errors and discount them accordingly.[2]

Similar to the listener-misperception approach to sound change, the speaker-oriented approach to sound change also assumes that phonologization is achieved at the lexical level. Lindblom (1990), for example, argues that speakers adaptively tune their performance along the H(yper)-H(ypo) continuum according to their estimates of the needs of the listener in that particular situation. While speakers and listeners dynamically adjust their production and perception to the communicative demands of the situation, sound change may occur when intelligibility demands are redundantly met or when the listeners focus their attention on the "how" (signal-dependent) mode of listening rather than on the "what" (signal-independent) mode of listening (Lindblom et al. 1995). New phonetic variants are accumulated during the 'how' mode of listening to the reservoir of perceptual memories and can serve as production models in future occasions. Like the listener misperception approach, systemic shifts require the accumulation of new variants. While Lindblom's speaker-oriented approach to sound change allows for the existence of listener misperception and reanalysis as a source of new variant acquisition, the main source of variation comes from the functional adaptive nature of speech communication, which gives rise to hyper- and hypo-articulation in speech. While the accumulation of perceptual mistakes or new perceptual variants cannot be ruled out, it remains a question how chance encounters with novel variants (mistakes or otherwise) can lead to systematic shifts that pervade the whole lexicon (i.e., the problem of implementation, see Bermúdez-Otero 2007). A stable and consistent source of novel variants must exist in order to sustain the introduction of a stable new variant from the listener-speaker perspective.

The goal of this paper is to offer a conception of phonologization from a perspective of sound change that puts the individual as the central locus of "change." Specifically, phonologization, as it is understood in this study, occurs whenever a speaker acquires a cognitively controlled sound pattern as the consequence of "permanent replication errors during grammar acquisition and grammar updating" (Bermúdez-Otero 2015: 12). This conception of phonologization offers a better framework for addressing the problems of precursor exaggeration and implementation mentioned above. The next section (Section 2) discusses what it means to conceptualize phonologization as an individual-level phenomenon. Section 3 illustrates this approach to phonologization with a case study of individual variation in vowel duration as a function of vowel height in Cantonese, arguing that a subset of Cantonese speakers have phonologized height-dependent vowel duration variation. This section is followed by a more general discussion in Section 4 regarding the proposed reconception in relation to past models of phonologization as well as ideas for future directions in phonologization research.

2 Phonologization from an individual-difference perspective

The conception of phonologization, as it is proposed here, presupposes a model of sound change that sees language change as a reflection of a difference between the grammars of individuals. That is, sound "change" is evident when one grammar, G_1, manifests

---

[2] It is worth noting that the idea that L2 speakers may induce sound change is not completely untenable. As Thomason & Kaufman (1988) pointed out, language shift can happen as a result of interference. That is, foreign elements may enter the language via L2 speakers who are treated by the L1 speakers as if they were L1 speakers. However, it is unclear whether such contact-induced changes should be conceptualized the same way as "endogenous" changes (i.e., language-internal sound shifts).

a sound or feature as *i*, while G_2, a "descendant" of G_1,[3] manifests the corresponding sound/feature of *i* as *j* in *corresponding* positions in *corresponding* lexical items (Hale 2007; Hale et al. 2015).[4] Phonologization within such a framework thus refers to an erstwhile phonetic precursor being acquired by a learner as an intended sound pattern under the cognitive control of the speaker. This type of "permanent replication errors" (Bermúdez-Otero 2015: 12) can happen during the process of child language acquisition, but it can also happen whenever grammar updating happens. The different phonologization outcomes often go unnoticed until there emerges a variation that is large enough to generate different descriptions in the coarse coinage of our shared language.[5]

   That individual variation exists in speech perception and production is a banal enough fact. Individual variability in speech production, for example, may come from differences in vocal tract physiology, particularly related to the nature of sexual dimorphism of the vocal tract (Vorperian et al. 2011), vocal tract size and shape (Peterson & Barney 1952), and/or behavioral/etiological factors (Sachs et al. 1973; Ohala 1994). But not all individual variability can be attributed to such physiological or non-linguistic factors. By locating phonologization at the individual level, the question of phonologization becomes a question about why certain individuals end up exhibiting unique perceptual and production strategies relative to other speakers within the communities of practice with whom they interact with (i.e., the proverbial "perturbation" to the community grammar; MacKenzie 2019), and how they may serve as innovators, though not necessarily as leaders, who can sustain the introduction of stable new phonetic variants in the community. Answers might reside in input biases, which stem from differences in personal experiences across speakers, or intake biases, which originate in variability across individuals in how they process the speech signals, perhaps as a result of differences in cognitive processing style (see Dediu & Moisik 2019 and MacKenzie 2019 for other potential sources of individual variability mentioned in this volume).

   Consider, for example, the cue reweighting path of sound change discussed above. The type of trading relationship between the disappearance of a voicing difference and the emergence of an f0 difference finds support in studies examining the malleability and variability of cue weighting. Francis et al. (2008), for example, showed that perceptual training can prompt listeners to adjust the reliance of one cue over another in their classification of speech sounds. From the production point of view, phonetic imitation studies (e.g., Babel 2012; Nielsen 2011) found that speakers could adjust their production patterns when exposed to production targets that differ from their production norms, although it is not clear if phonetic imitation in one phonetic dimension can lead to adjustment in another dimension. Crucial from an individual-difference perspective are recent studies, including investigations regarding the relationship between VOT and f0 in stop voicing perception and production, which documented extensive individual variation in

---

[3] The notion of "descendant" is to be understood in very general terms. That is, if G_2 is acquired or updated based on the outputs of G_1, G_2 is a "descendant" of G_1.

[4] The notion of grammar is broadly construed here. Crucial for the model of phonologization below, any cognitively controlled patterns, gradient or otherwise, are assumed to be part of an individual's phonological knowledge and may thus undego "change".

[5] There are some reports of sound change over the course of an individual's life span (Harrington et al. 2000; Harrington 2006; Sankoff & Blondeau 2007), but as noted in Joseph (2011), it is generally difficult to know if the observed fluctuations are changes in an individual's performance in a very circumscribed and controlled set of circumstances or perhaps even changes in one's perception of how one could or should speak in a certain context, since the analyst generally does not have access to information about how an individual speaks in all other contexts. If the individual's linguistic competence already encompasses a wide range of ways of speaking, including those observed over the person's life time, the observed changes over the life span should not be taken as evidence of sound change *per se*.

cue weighting for a variety of cues (e.g., Hazan & Rosen 1991; Escudero & Boersma 2004; Idemaru et al. 2012; Shultz et al. 2012; Schertz et al. 2015; Kong & Edwards 2016; Kapnoula et al. 2017; Clayards 2018a) and such variable cue weight settings across individuals are stable over time (Idemaru et al. 2012; Schertz et al. 2015). These findings suggest that the so-called "exaggeration" of phonetic precursor is a preexisting condition, as it were. That is, from the perspective of the individual who relies more on f0 than VOT, it does not make sense to speak of "exaggeration" *per se* since that individual's cue weight setting, for all intents and purposes, has always been set that way. No change has taken place within this individual, modulo the regular grammar updates from daily experience. That is, this person's perceptual and production grammars might very well have always given f0 more perceptual and production weights than other members of her speech community. An independent observer might discern the existence of a "change" when this individual's perceptual and production grammars (i.e., weight settings) are compared to those of her peers. To be sure, there remains a question of why this person has the cue weights set so differently than others. As noted earlier, one possible explanation might stem from individual differences in past experience (e.g., Schertz et al. 2015; Lev-Ari & Peperkamp 2016). But of particular interest here are recent reports that such individual variability in cue weighting may stem from individual-specific cognitive traits (Clayards 2018a; Kong & Edwards 2016; Kapnoula et al. 2017). For example, individuals who are more gradient in categorizing a continuum of acoustic signals are more likely to utilize secondary cues in speech perception (e.g., Kong & Edwards 2016; Kapnoula et al. 2017; Ou et al. In press), which might in turn be related to individual difference in cue processing strategy (Ou et al. In press). Variation in categorization gradience of speech might also stem from individual variation in speech signal processing at the neuro-level (Ou & Yu 2019; Ou & Yu Under review). Individual variation in executive functions has also been implicated in individual variation in cue reliance (Kong & Edwards 2016; Kong & Lee 2017).

A similar method can be taken within the listener misperception/reanalysis approach to sound change. Recall that this approach focuses on the ambiguous nature of the speech signal, which gives rise to opportunities for listeners to parse the signal in ways that differ from their interlocutor's. Advocates of this approach often assume that listeners are generally good at factoring out (i.e., perceptually compensating for) the source of variation during processing (e.g., the presence of vowel nasalization is seen as an artifact of anticipatory coarticulation from a following nasal or the lowered spectral frequencies of a sibilant is attributed to the presence of a following rounded vowel, etc.). But occasional misparsing might lead to the creation of novel variants. The perspective taken in this work would recast this scenario from an individual-difference perspective. As recent studies have shown, listeners react to coarticulatory information in the signal in diverse ways (e.g., Repp 1981; Beddor 2009; Yu & Lee 2014; Zellou 2017; Beddor et al. 2018; Yu 2019). Some individuals would indeed exhibit behaviors that suggest perceptual compensation for coarticulation, while others would exhibit more veridical perception and not take into account the presence of a potential coarticulatory source. Still others exhibit behaviors that are in between the two extremes. Such individual variation has been found to be consistent, albeit only moderately, across tasks (Yu & Lee 2014). Of particular interest here are recent reports that found variation in perceptual strategies for dealing with coarticulatory information (Yu 2010) as well as that the corresponding production of coarticulated speech (Yu 2016) may be related to differences in cognitive processing styles.

These findings thus not only lend support to an individual-specific conception of phonologization, but also to the importance of integrating individual-difference methodologies into sound change research. To be sure, there is still much to be learned about the origins of individual variation in intake biases and how they shape the acquisition and updating of individual phonological knowledge. However, without a proper framework for understanding phonolologization from an individual-difference perspective, certain questions might remain difficult to address. For example, how do gradient sound patterns acquire categorical characteristics? As noted earlier, recent research has identified significant individual variability in speech categorization. That is, for a given speech signal, some individuals are more inclined to classify the signal as part of one category and not another, while others might map the signal with less certainty, preferring instead to treat the signal more veridically. Are individuals who classify speech more categorically and with more certainty more likely to develop categorical patterns, and are more gradient categorizers more likely to retain the gradient characteristics of a sound pattern? These questions are difficult to conceptualize and address within a population-average model of phonologization, or, indeed, cognition in general.

From the present perspective then, the traditional appeal to channel biases (i.e. the type of articulatory, acoustic, auditory an perceptual constraints inherent to the vocal tract, along with the auditory and perceptual apparatus as alluded to above) is, perhaps paradoxically, not the emphasis here, as such biases are presumably always present and not unique to any particular individual.[6] To be sure, identifying channel biases remains an important pursuit as such biases are the bases of variability in the speech signal. But as an explanation of sound change, and of phonologization in particular, the presence of channel biases is necessary, but not sufficient. One must be able to identify the individuals who treat channel biases not as noise, but as linguistically relevant information intended by the speaker, in order to advance the discussion. Specifically, an individual who acquired a phonetic variation as cognitively-controlled is an individual who has phonologized that sound pattern. Rather than conceptualizing the difference between intrinsic/automatic vs. controlled phonetic variations as a difference between language communities or two historical stages of a language, it is more fruitful to consider how individuals within the same community might vary in how much control they have over a phonetic variation. This conception of phonologization allows us to see the problems of precursor exaggeration and implementation in a completely different light. The cue-reweighting path to sound change does not begin with a stage where individuals within a given speech community exhibit intrinsic phonetic variation at time k transitioning to a stage where individuals within the same community exaggerate the precursor at some later point, $k+n$, in time. Rather, the individual-difference conception of phonologization contends that, at any given point in time, the potentiality of someone acquiring a phonetic variation as a speaker-controlled pattern exists, provided that the pattern supported by the phonetic variation is learnable. The "precursor exaggeration problem" becomes a question about the nature of grammar acquisition.

The individual-difference perspective of phonologization also deemphasizes the reliance of occasional accumulation of deviant exemplars as a way to actuate systematic change. For a given phonetic variation in the signal (i.e., the precursor), if an individual acquires

---

[6] This presupposition is itself an idealization, to be sure, since small anatomical differences across individuals may result in different articulatory strategies that can affect the acoustic output (see review, e.g., in Dediu et al. 2017) and influence sound change (e.g., Baker et al. 2011; Dediu & Moisik 2019; Smith et al. 2019). Thus, the point here is not to deny the possible influence of individual differences in channel bias on sound change, but to steer the discussion away from population-level thinking that underlies earlier work on the effects of channel bias on sound change.

it as controlled variation, thus part of the individual's grammatical system, all forms that satisfy the condition of the process would be subjected to the same alternation. For example, if an individual learns to weigh the f0 cue as more important than the VOT cue, under the assumption of the uniformity constraint (Chodroff & Wilson 2017; 2018; Chodroff et al. 2019), all stops in the corresponding context should receive similar weighting. If nasality in VN sequences is analyzed by the individual as prosodic, then the same analysis should apply to VN sequences found in any lexical items.

To summarize the discussion thus far, an individual-difference approach to phonologization has the following characteristics:

- The proposal
    - The phonologization of phonetic precursor X occurs when an individual internalizes the precursor as a cognitive-controlled pattern, X′.
- Corollaries
    - Understanding phonologization requires understanding if someone has internalized a variation as a controlled pattern or not.
        * What are the best methods for uncovering controlled phonetic variation?
    - Whenever a phonetic variation X exhibits individual variability, one should investigate how members of the community internalize X; some might internalize X as part of the phonological grammar while others might not.
        * What factors contribute to individuals internalizing X differently? Past linguistic experiences? Intrinsic neuro-/cognitive differences?
        * What is the distribution of the individuals who have phonologized X? Are they in the majority? How are they situated within the social semiotic landscape (Eckert 2019)?
    - Stable implementation of a phonologized pattern stems from the fact that the pattern is part of the grammar of the innovating individual.

Before expanding further on the implications of this perspective on phonologization, the next section offers an illustration on how one might approach a case of individual variability in phonetic variation from the perspective laid out thus far.

3 Conceptualizing individual variability in intrinsic vowel duration

The phonetic variation under consideration is the so-called intrinsic vowel duration (IVD), which refers to the positive association between vowel height and duration. IVD has been observed across a wide variety of the world's languages, such as Catalan (Solé & Ohala 2010), English (Heffner 1937; House & Fairbanks 1953; Peterson & Lehiste 1960; Scharf 1962; Solé & Ohala 2010; Toivonen et al. 2015), German (Fischer-Jorgensen 1940; Maack 1949), Japanese (Solé & Ohala 2010), Inari Sammi (Äimä 1918; Stone 2014), Swedish (Toivonen et al. 2015), Thai (Abramson 1962), and Spanish (Navarro Tomás 1916), yet investigations into the question of individual variation have been scarce (though see Solé & Ohala 2010; Yu et al. 2014).

The explanation for this positive correlation has been a matter of debate. Some scholars have argued for a biomechanical interpretation, suggesting that the motor commands for timing might be the same across vowels, but certain vowels would take longer to produce because of the extra time it takes for the jaw to open (e.g., Fischer-Jorgensen 1964; Lindblom 1967). However, duration differences persisted even in experiments where the jaw position was fixed (Nooteboom & Slis 1970; Smith 1987). Other studies have suggested that the durational differences might be phonologized. Lisker (1974), for ex-

ample, pointed out that the often long steady-state formant structures of low vowels are inconsistent with the idea that low vowels are longer because of the time it takes the jaw to move, as one might expect the onset and offset formant movements toward low vowels to be longer than the steady-sate portion. Tauberer & Evanini (2009) found that duration does not increase as vowels are lowered in language change. Solé & Ohala (2010) examined the effects of speaking rates on vowel duration in Japanese, Catalan, and English, and found that duration differences expand with slower speaking rates among vowels of different heights in Catalan and English and between long and short vowels in Japanese, but the duration differences remained constant between vowels of different heights in Japanese. They interpreted the results as supporting the conclusion that the correlation between vowel duration and vowel height in Catalan and English is controlled, and the Japanese pattern of IVD is due to biomechanical reasons (see below for more discussion). More recently, Toivonen et al. (2015) examined the between- and within-vowel category variation in duration in English and Swedish and found that high vowels are indeed shorter than low vowels, but higher instances of the vowel [ɪ] are not shorter than lower instances of [ɪ], suggesting that a purely biomechanical account of IVD is untenable.

From the perspective of sound change, IVD is intriguing as there are few reported cases of sound patterns that are claimed to be reflexes of such phonetic variation. One such example is found in Dutch, where original long /i, y, u/ have phonologized as short /i, y, u/ and have merged in their quantity with /ɪ, ʏ/ (Gussenhoven 2004). By focusing on how IVD might vary across individuals, it might shed light on how pervasive and robust this variation is within and across individuals within a speech community, which in turn might offer information regarding the likelihood of IVD propagating as a sound change across a community. In what follows, we examined individual variability in the production of vowel duration as a function of vowel height in Hong Kong Cantonese. We then interpreted the findings from the perspective of phonologization as advocated above.

3.1 Materials

The recordings analyzed in this study were originally obtained as part of a larger study of Cantonese phonetics and phonology. Our analysis focused on syllables with high and mid vowels only. In particular, eight target words were targeted: [syˉ] 'book', [sy↓] 'potato', [soˉ] 'comb', [so↓] 'silly', [siˉ] 'poem', [si↓] 'time', [seˉ] 'a little bit', [se↓] 'snake'. The syllable with /a/ was not included in this analysis because the syllable /sa55/ does not have a T21 counterpart. Cantonese has a six-way tonal contrast: high level [ˉ], mid-rising [ˊ], mid-level [˧], low falling [↓], low-rising [ˏ], and low-level [˩]. The high level [ˉ] is in free variation with a high falling [ˋ] tone. The low falling [↓] is often realized and transcribed as [˩]. For ease of reference, [ˉ] will be referred subsequently as T55 and [↓] as T21.

3.2 Participants

Ninety-five native speakers of Hong Kong Cantonese (twenty-eight males; median age 19, with age ranges from 17 to 26) with no reported history of speech, language, or hearing problems were recorded in Hong Kong in a quiet room as part of a larger study of Cantonese phonetics and phonology. All were undergraduates at a university in Hong Kong. Each subject received a nominal fee or course credit for participating in the study.

3.3 Procedure

Each participant was digitally recorded in a quiet room individually at a sampling rate of 44,100 Hz reading three blocks of the target stimuli, presented in one of two pseudo-randomized lists of target words in the carrier sentence, [ŋɔ˧ tʊk˥˧] _ [pei˧ nei˧ tʰeŋ˥] 'I read _ for you to hear.' A total of twenty-four target stimuli were analyzed for each participant. The stimuli were presented in traditional Chinese characters. All subjects also completed an online survey that included questions regarding the subjects' personal demographics and personality traits.

3.4 Analysis

Vowel duration was modeled using linear mixed-effects regression fitted in R, using the lmer() function from the lme4 package (Bates et al. 2011). Vowel duration that are 2.5 standard deviations from the mean were not included in the analysis, which amounted to less than 1.4% data loss.

The regression model tested for the effects of trial order (TRIAL; 1-44), vowel HEIGHT ([i y] vs. [e o]), and TONE (T55 vs. T21). Both HEIGHT and TONE were sum-coded. Trial order was centered and z-scored. While the focus of this case study is on the effect of vowel height on duration, TONE is included to controlled for potential variation in duration as a function of TONE (e.g., Gandour 1977; Kong 1987). The model also included by-subject random intercepts to allow for subject-specific variation in vowel duration. Given that the inclusion of by-subject random slopes for HEIGHT and TONE significantly improves model likelihood, both independently and together, they were included in the final model to allow for by-subject variability in the effects of vowel height as well as tones on the vowel duration. Models with by-subject random slopes for the interaction between HEIGHT and TONE did not converge and were therefore not included in the final analysis. The model formula in lme4 style for vowel duration is DURATION ~ TRIAL + HEIGHT * TONE + (1 + HEIGHT + TONE |SUBJECT).

The residuals of the initial fit were examined and found to deviate strongly from normality. As a result, residuals that were more than 2.5 standard deviations from the mean were trimmed, which amounted to no more than 2.5% of the data, and the model was refitted to the trimmed data set. The new model had a residual distribution much closer to normality, and it is the refitted models that are reported below.

3.5 Group-level analysis

Table 1 summarizes the regression model for vowel duration. Figure 1 shows the average vowel duration in syllables with different vowels and tones. Qualitatively, high vowels are shorter than mid vowels and there is no obvious tonal difference in duration except when the vowel is /o/. The regression model confirmed these observations. There is a significant difference in vowel duration between high and mid vowels (β = -11.97, t = -19.64, $p < 0.001$) such that the high vowels are shorter than the mid ones. There was also a significant effect of TRIAL (β = -5.22, t = -10.26, $p < 0.001$), suggesting that speakers generally spoke faster as the experiment progressed.

While TONE is not a significant predictor of vowel duration, there is a significant interaction between HEIGHT and TONE. That is, the effect of tone on vowel duration is modulated by HEIGHT (HEIGHT:TONE: β = 1.57, t = 3.11, $p < 0.002$) such that the tone-based difference in vowel duration is only found among the mid vowels. Since this discussion

focuses on the interaction between vowel height and duration, the tonal effects on vowel duration will not be discussed further below.

Table 1: Regression model results for vowel duration. $* = p < 0.05$, $** = p < 0.01$, $*** = p < 0.001$. The p-values were obtained using normal approximation, which assumes that the t distribution converges to the z distribution as degrees of freedom increase (see Mirman 2014 for details).

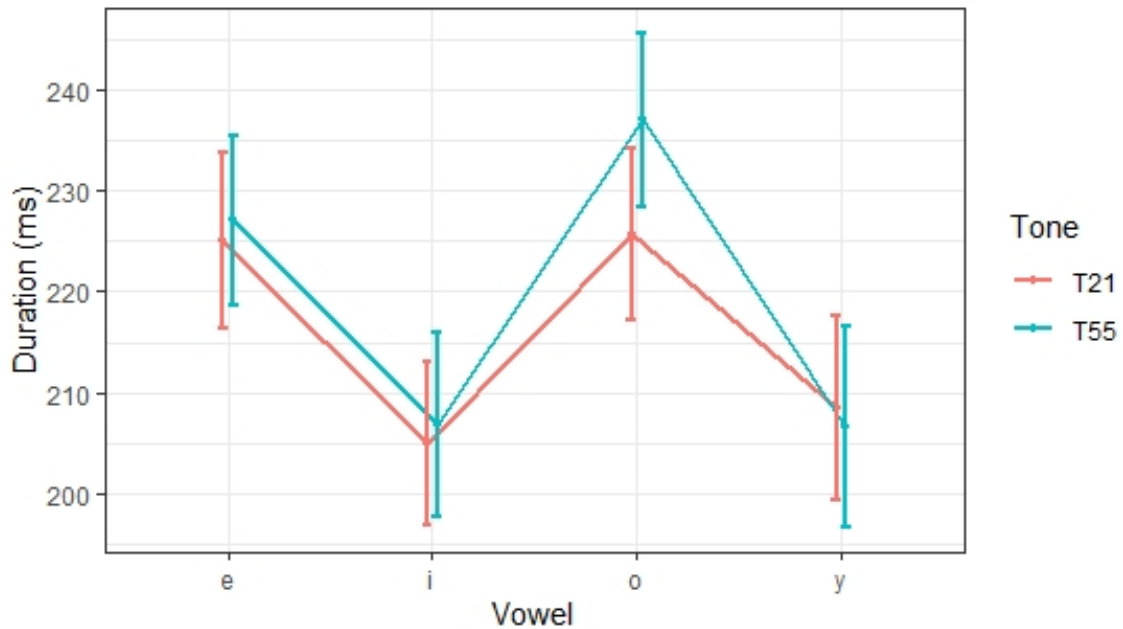|	Coef (SE)	t-value
INTERCEPT	217.16 (4.20)	51.65 ***
TRIAL	-5.22 (0.51)	-10.26 ***
HEIGHT	-11.97 (0.61)	-19.65 ***
TONE	-1.12 (0.65)	-1.73
HEIGHT:TONE	1.57 (0.50)	3.11 **

Figure 1: Average vowel duration for sV syllables with different vowels and tones. The error bars indicate 95% confidence intervals.

3.6 Individual-level analysis

The group-level analysis shows that this cohort of Cantonese speakers exhibited vowel duration differences based on vowel height. The higher the vowel, the shorter the duration. The fact that the inclusion of by-subject random slopes for VOWEL independently and significantly improves model-likelihood, suggests that there is marked individual variability in the magnitudes of the effects of vowel height. Figure 2 shows the nature of individual variation in vowel duration. Individuals with error bars above and below zero can be interpreted as having duration differences that are significantly different from the group mean. To illustrate this variation further, the left panel of Figure 3 features three individuals (participants 42, 73, 80) who have random slope values for HEIGHT that are above zero, suggesting a reduced or null IVD effect. The right panel of the same figure

features individuals whose random slopes for HEIGHT are below zero (participants 10, 36, 50), suggesting a great vowel duration difference across vowel heights.
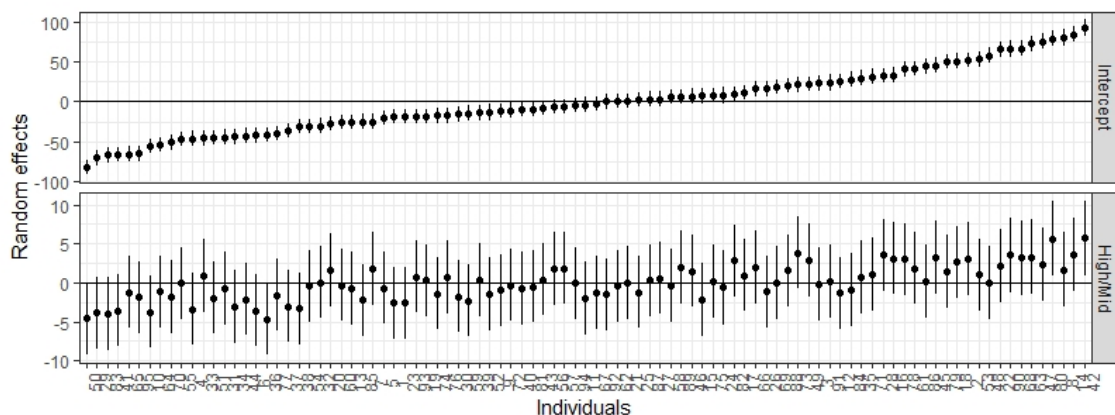

Figure 2: "Caterpillar plots" for the conditional modes of the by-subject random intercept and the by-subject random slopes for HEIGHT for vowel duration across 95 participants. The participants are ordered by the conditional modes of the by-subject intercepts.
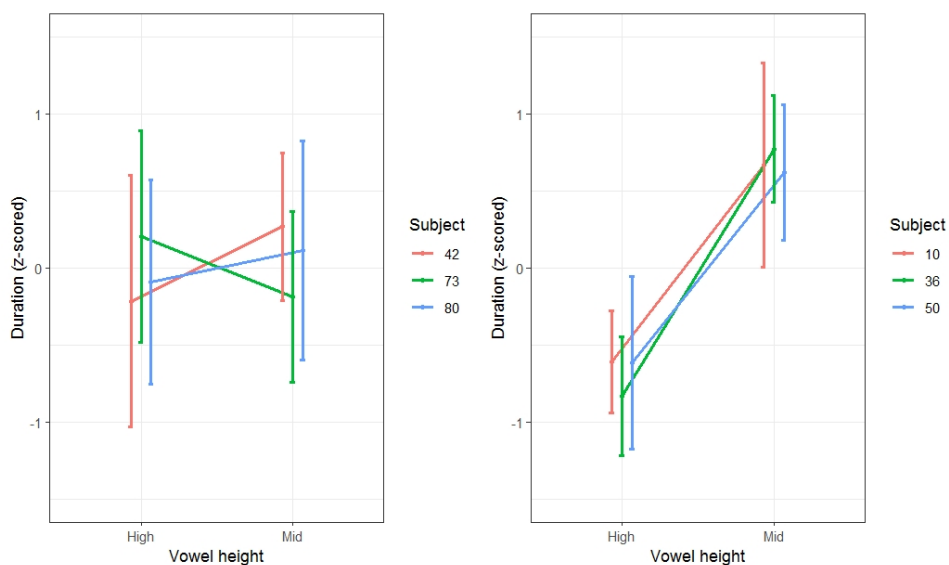

Figure 3: Vowel duration across vowel heights for subjects 42, 73, 80 (random slopes above zero) and subjects 10, 36, 50 (random slopes below zero). Duration values are z-scored relative to the individual mean and standard deviation. Error bars present the 95% confidence intervals.

The fact that individuals vary in the way vowel height affects vowel duration suggests that the effect is unlikely to be biomechanical in origin, at least not entirely, for if the variability was due to purely biomechanical reasons, all participants should exhibit similar patterns. One potential source of the observed individual differences might have to do with some speakers internalizing the vowel duration differences as a controlled phonetic variation, while others have not. The controlled nature of IVD can be further ascertained by examining how IVD might vary according to global duration conditions.

In a series of studies (Solé 1995; 2007; Solé & Ohala 2010), Maria-Josep Solé proposes that the controlled nature of a phonetic variation can be diagnosed by considering how a phonetic pattern is manifested across different speaking rates or stress conditions. In her investigation of stop voicing contrasts in English and Catalan, for example, Solé (2007) argues that voice onset time (VOT) is a language-specific property of English used to signal voicing contrasts as evidenced by the fact that English speakers adjust their VOT as a function of speaking rate variation. The short positive VOT values found in Catalan, on the other hand, do not change as a function of speaking rate variation and were seen as reflecting the aerodynamic link between the stop release and voicing, and thus not a controlled phonetic cue. More recently, Chodroff and colleagues (Chodroff & Wilson 2017; 2018; Chodroff et al. 2019) investigated the positive VOT for six word-initial stops in English and found that, while mean VOT for each stop differed considerably across talkers, there was a strong linear relation among the means across places of articulation, suggesting the presence of a uniformity constraint on the talker-specific realization of a phonetic property, such as glottal spreading.

Concerning intrinsic vowel duration, Solé & Ohala (2010) suggest that if height-dependent duration differences were controlled, such differences would be expanded if speakers aim to maintain a relatively constant perceptual difference across speaking rates. Unlike the treatment of VOT discussed above, Solé & Ohala (2010) proposed, though without argument, that one might expect a constant vowel duration difference between slow and fast speech rates or a slightly smaller difference in faster rates if vowel duration differences were due to biomechanical factors. It is worth noting that if low vowels were longer because of the time it takes the jaw to move, one might expect vowel duration difference would also reduce or evaporate in a slower speaking rate (cf. Lisker 1974). Solé & Ohala (2010) found that the controlled vowel duration hypothesis is borne out in Japanese, where the phonemic vowel length difference in Japanese interacts with speaking rate in a way that the duration difference between long and short vowels is larger at slower rates than in faster rates. The difference in duration between high, mid, and low vowels within the same vowel length category remain stable across different speaking rates. They conclude that the short-long difference in Japanese is under control of the speaker, but the vowel height differences reflect a mechanical effect. Similar effects of duration ratio maintenance were observed in the case of English vowel height and tenseness contrasts. That is, differences in duration across different vowel heights and tenseness are larger in slower speaking rates than faster ones. This situation is considerably more complicated in Catalan, however. In their investigation of the duration properties of Catalan /i, e, ɛ, a/, Solé & Ohala (2010) found that speakers scale up the duration differences between (mid)close, /e i/, and (mid) open vowels, /ɛ a/, at slower speaking rates, as opposed to differences between vowels within the same close or open category. That is, the difference between /i/ and /e/ and between /ɛ/ and /a/ appear to be stable across speaking rates. They conclude that Catalan speakers are manipulating duration differences to cue high vs. low vowels, but did not offer an explanation in terms of why the duration difference within the open and close categories shows a stable duration difference.

The predictions regarding the relationship between speaking and IVD can be summarized as follows:

- If intrinsic vowel duration were biomechanical in origin...
    - Vowel duration difference should stay roughly constant across speaking rates across speakers (Solé & Ohala 2010) but the difference should disappear when speaking is slow (Lisker 1974).
- If intrinsic vowel duration were controlled...

- Vowel duration difference might expand or contract as a function of speaking rate in order to maintain a constant duration ratio between vowels (Solé & Ohala 2010).
- Vowel duration difference might remain consistent across speaking rates, but the vowel duration difference is talker-specific (Solé 2007; Chodroff & Wilson 2017; 2018; Chodroff et al. 2019).

While the present data set was not originally designed to examine the effect of speaking rates on IVD, the effects of speaking rates on vowel duration can nonetheless be examined by focusing on the interaction between Trial and Vowel. Recall that the regression model for the production data revealed a significant effect of Trial, indicating that the participants as a group were speaking faster as the task progressed from trial to trial. To explore the interaction between between speaking rate, as indexed by the significant trial order effect, and the effect of vowel height on vowel duration, individual regression models were fitted for each subject's vowel duration values. The regression models were fitted using the ddply() function in the plyr package (Wickham 2012). To simplify the analysis, the regression models included as predictors only Trial and Height, and the interaction between these two factors. The model formula is Duration ∼ Trial * Height. Figure 4 illustrates the duration differences between vowel height over the course of the experiment; the figure only shows the twenty-nine of the ninety-five participants who exhibited significant Trial and Height effects at $p < 0.05$ level. Out of these twenty-nine participants, only two exhibited a significant interaction between Trial and Height. Subject 51 exhibits the type of duration ratio maintenance effect Solé & Ohala (2010) referred to (Trial x Height: $\beta = 7.86$, t = 2,52, $p < 0.05$). That is, at a slower speaking rate (earlier in the experiment), the duration difference between vowels of different heights is larger than at a faster speaking rate (toward the end of the experiment). Subject 41 exhibits the opposite effects (Trial x Height: $\beta = -5.24$, t = -2,26, $p < 0.05$). That is, the height-dependent duration difference is largest when the speaking rate is fastest. While some of the other subjects exhibit trends that seemed suggestive of a Trial x Height interaction (e.g., 15 and 52 are similar to Subject 51, while 26, 23, 34, and 67 are similar to 41), those interactions did not reach significance, suggesting that the duration difference between high and mid vowels generally stays constant regardless of speaking rate. The findings from the individual-level analysis are echoed at the group level. That is, the inclusion of an interaction between Trial and Height did not improve significantly model likelihood from a model without such an interaction. Such a result is consistent with the idea that the height-dependent vowel duration differences did not vary across trials, even though the overall duration of the vowels decreased trial after trial.

Figure 4 suggests that, for virtually all the speakers who exhibit a Trial effect, the duration difference between vowels of different height is maintained, which is consistent with the idea that a uniformity constraint is present to keep vowels of different heights durationally distinct. But the way such a distinction is maintained might differ from individual to individual. Rather than maintaining the difference in terms of duration ratios between vowels like Subject 51, many speakers maintain the distinction with a constant acoustic difference even when the baseline duration of each vowel varies across individuals, similar to the uniformity effect observed for VOT across stops of different places of articulation (e.g., Chodroff & Wilson 2017). The patterns of subjects 10, 23, 26, 34, and 31 are intriguing since they suggest that a greater duration difference is found when speaking rate increases. Such results are in line with the biomechanical explanation laid out above. That is, at a slower speaking rate (i.e., at the beginning of the recording),

the vowel duration difference is small. In particular, with the exception of Subject 10, high vowels are generally pronounced shorter at a faster speaking rate, but the duration of the mid vowels remained largely stable. Finally, as already noted above, the fact that individuals vary in terms of the magnitude of the vowel duration differences, with some individuals not showing much of a difference at all (i.e., the 46 individuals who did not show any HEIGHT effect), suggests that the biomechanical contingencies that are assumed to underlie IVD do not always give rise to detectable vowel duration differences.
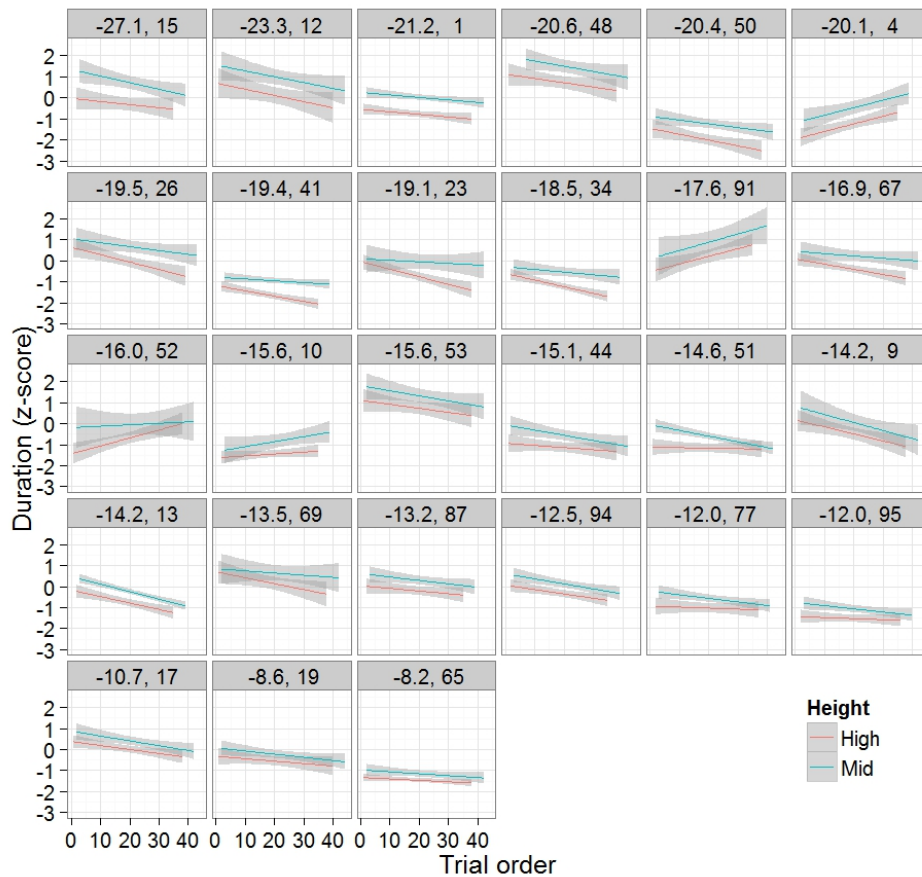

Figure 4: Each panel shows the vowel duration patterns as a function of vowel height across trials by one subject. Panels are ordered by the size of the HEIGHT estimate, as indicated in the first value in the facet header of each panel; the second value indicates subject number. Regression lines indicate the effects of trial order on vowel duration across high and mid vowels. The ribbons present the 95% confidence intervals. Data shown here are from twenty-nine participants who exhibit significant TRIAL and HEIGHT effects at the $p < 0.05$ level. Duration values were z-scored relative to the individual mean and standard deviation.

Another issue related to the approach toward investigating individual variation taken here, as noted by one of the reviewers, is whether the individual patterns reported are genuine inherent properties of individuals or whether they might reflect false positives that are the results of sampling a noisy process. To this end, we use a holdout method of cross-validation to examine the extent to which the estimates for the by-subject random intercepts and slopes fluctuate after resampling. The cross-validation procedure was as follows: the regression model described above for the production data was applied to two

randomly selected halves of the sample; the two samples were complimentary to each other. To assess how much fluctuation would result from resampling, correlation analyses were conducted between the by-subject random slope estimates based on the whole sample and those based on the two randomly selected halves. Figure 5, which summarizes the results of the correlation analyses, shows that estimates across resamplings are sufficiently similar to those obtained from the whole sample (i.e., r above 0.77 with $p < 0.001$). The finding of strong correlations across resamplings is consistent with the idea that individual patterns reported here are genuine inherent properties of the individuals.
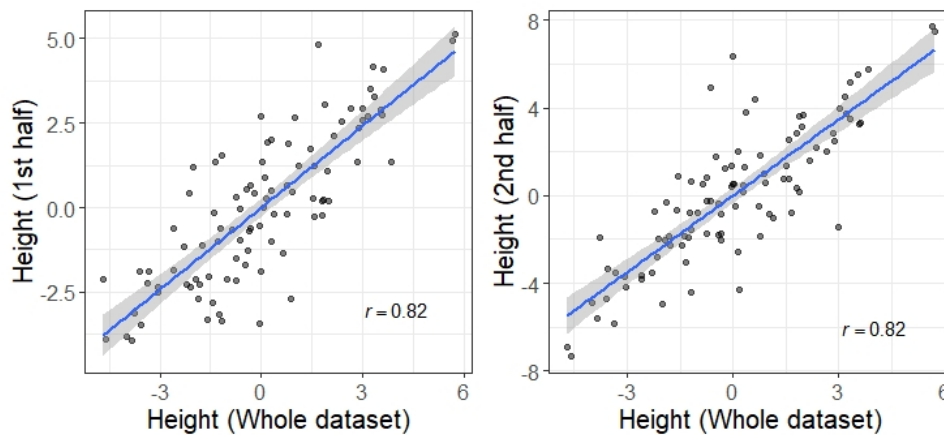

Figure 5: Scatterplots showing random slope estimates for Hᴇɪɢʜᴛ based on the whole data set and the two randomly selected halves of the data. All correlations are significant at the $p < 0.001$ level.

Related to the question of the nature of individual variation is the relative importance of individual-level factors and the fixed, group-level, factors. To this end, we followed Nakagawa & Schielzeth (2013) and Johnson (2014) and calculated the marginal R^2, which describes the proportion of variance explained by the fixed factors alone, and the conditional R^2, which describes the variance explained by the combination of the fixed and random factors. The difference between the two R^2 values is the proportion of variance explained by the random (i.e., individual-level) factors. In the case of the regression model for the production data, for example, the marginal R^2 is 0.066 and the conditional R^2 is 0.768, suggesting that the variance explained by the individual-level factors is more than 10 times that of the fixed factors. To be sure, much of the variance is explained by cross-individual differences in the general duration profile. Table 2 summarizes the marginal and conditional R^2 values for a series of regression models with varying degrees of model complexity. The first model, M0a, includes no fixed factors and only by-subject random intercepts. Given that the marginal R^2 is zero and the conditional R^2 is 0.638, the proportion of variance explained by the random intercepts is 63.8%. With the fixed factors added, the proportion of variance explained by the random intercepts changes slightly to 64%. Crucially, the increases in proportion of variance explained by the inclusion of by-subject random slopes for Tᴏɴᴇ, Hᴇɪɢʜᴛ, or the combination of the two, ranges from 0.9% to 1.4%. Compared to the proportion of variance explained by the fixed effects, which hovers around 6.2%, the strength of the individual-level effect is sizable (around a 1/6 to almost a quarter of that of the fixed factors).

Table 2: Comparisons of the proportions of variance accounted for by the marginal and conditional R^2 values across models with different complexities.

	Model formula	mR^2	cR^2	Δ
M0a	(1|Subject)	0	63.8%	63.8%
M0b	Trial + Height * Tone	5%	5%	0
M0c	Trial + (1|Subject)	0.5%	64.3%	63.8%
M1	Trial + Height * Tone + (1|Subject)	6.2%	70.2%	64%
M2a	Trial + Height * Tone + (1 + Tone|Subject)	6.2%	70.8%	64.6%
M2b	Trial + Height * Tone + (1 + Height|Subject)	6.1%	70.5%	64.4%
M3	Trial + Height * Tone + (1 + Height + Tone|Subject)	6.2%	71.1%	64.9%

3.7 Discussion

The above case study shows that, from the perspective of the group-level analysis, Cantonese speakers maintained a constant duration difference between high and mid vowels throughout the experiment, despite the increasing speaking rate across trials. This overall scenario, we argue, suggests that IVD in Cantonese is a controlled phonetic pattern, following similar arguments laid out in previous studies (Solé 2007; Chodroff & Wilson 2017; 2018; Chodroff et al. 2019). To be sure, this interpretation of our results might seem, at first glance, inconsistent with the conclusion Solé & Ohala (2010) drew in their study of Japanese, Catalan, and English, where they interpreted a constant duration difference between vowels of different height (regardless of speaking rate) as evidence for a physiological explanation. There are reasons to treat their conclusion with caution, however. To begin with, unlike their study, which focused on differences across groups of speakers, the IVD observed in our study is talker-specific. A mechanical explanation is unable to account for such individual variability, at least not without appealing to significant individual variability in vocal tract physiology. Moreover, as already reviewed above, a constant duration difference across speaking rates has been used as evidence for speaker control (Solé 2007). Thus it is unclear why such an interpretation is not suitable in the case of IVD. Moreover, while Solé & Ohala (2010) concluded that the magnitude of the vowel-dependent duration difference between long and short vowels in Japanese is proportional to the duration of the vowels across speaking rates, they did find significant differences in vowel duration ratio between fast and slow rates for certain vowel pairs in particular speakers. Given that their study involves only three speakers, and their general conclusion is based on a lack of a significant interaction between vowel height and speaking rate as supporting the null hypothesis, it is not clear that a constant duration ratio is the only correct diagnostic for speaker control. More importantly, from the individual-difference perspective advocated in this paper, the fact that someone like subject 51 exist (i.e. someone who maintains a constant duration ratio) suggests that the potential for IVD to be phonologized in this way is there. It should also be noted also that Cantonese has a vowel length contrast, albeit only in low vowels in closed syllables; there is also a difference in vowel quality (e.g., [fɐt] 'to punish' vs. [faːt] 'to rise, to become rich'). The fact that Cantonese speakers do not employ a duration ratio maintenance strategy in different speaking rates might be attributed to a ceiling effect. That is, a short low vowel can only be expanded so much before it is in danger of being confused for a long vowel. To the extent that there is a uniformity constraint over duration differences across vowels (Wilson & Chodroff 2017), a constant duration strategy might be the preferred strategy, at least for some Cantonese speakers, to maintain vowel duration differences.

This hypothesis cannot be directly tested here since only high and mid vowels in open syllables were analyzed. However, it is worth noting that the participants in the study did record low vowels as well as syllables that were closed. A quick examination of the /sa55/ syllables did find them to be longer than the syllables with mid and high vowels.

Intrinsic vowel duration phonologized: The individual-level analysis reveals that Cantonese speakers do not necessarily share the same height-dependent duration differences. While some maintain a constant duration difference between vowels across speaking rates, others show less durational differences in slower rates, while still others show more durational differences in slower rates. This state of affairs, we argue, suggests that Cantonese speakers might have internalized duration difference between vowels differently. From the perspective of the phonologization approach advocated in this study, individual variation in phonologization is to be expected. As hypothesized earlier, individuals within a speech community might vary in how they process and experience similar speech inputs. In the case of IVD in Cantonese, some individuals have apparently internalized the duration difference as a phonological pattern and therefore maintain the duration difference across trials. Others appeared to have discounted the pattern (i.e., internalizing the vowels as having similar duration) and produced either no discernible systematic differences in duration between vowels or small differences attributable to biomechanical contingencies. The reason for such differences in production behavior is unclear at this point. Previous studies have found significant individual variability in perceptual compensation for IVD effects (Yu et al. 2014). Individual differences in IVD variation might stem from individual variation in perceptual responses to the speech inputs. For example, individuals who compensated in their perceptual responses to the production asymmetry might have different production patterns than individuals who do not compensate. It is worth noting that whether or not there is a perception-production link is not directly relevant to the question of phonologization at hand. From the perspective of sound change laid out above, IVD is phonologized by individuals who exhibit systematic and stable vowel-height dependent duration variation in production.

Sound change in progress? Given that there exist individuals within the Cantonese-speaking community who have internalized IVD as a controlled pattern (i.e., IVD has been phonologized), an obvious question is whether the observed height-dependent vowel duration differences indicate a sound change in progress in Cantonese. This question can be stated even more generally. That is, whenever a phonetic variation is found to be cognitively controlled by some individuals within a community, is the phonologized pattern a sound change in progress? The answer to this question depends on how "sound change in progress" is defined. For the individuals who have acquired the pattern as part of the sound system of his/her language, relative to others who have not, a "sound change" has taken place. There is no sense ins talking about the change "progressing" at that level. From the larger community point of view, however, the question becomes a matter of sound change propagation. That is, the conceptual distinction between the *phonologization* of a variant within an individual and the *propagation* of a variant across individuals is key. As many scholars have noted (Ohala 1981; Milroy & Milroy 1985; Croft 2000; Eckert 2019), the propagation question must be addressed at the social level. For example, how is an individual who exhibits the innovative variants embedded socially within the speech community? What social meaning can be associated with the innovative variant? The discovery of speakers who have internalized a particular phonetic variation suggests that the potential for the innovation to propagate exists in the community. Whether an individual's "innovation" would be imitated and spread to the rest of the community would depend on many other contingencies (Baker et al. 2011). To be sure, as Harring-

ton et al. (2018) recently noted, the conditions that give rise to sound change and those that are responsible for its spread throughout the community might be more "artificial" (708) than real. Yu (2013), for example, argues that the cognitive processing style that contributes to an individual treating the speech signal differently from others might also contribute to how s/he interacts with others in the social world. An individual-difference perspective on phonologization provides the necessary theoretical framework to examine such a connection.

Implication for sound change typology: The fact that IVD can be phonologized (i.e., cognitively controlled) raises a curious question regarding sound change typology. As noted above, there is a dearth of sound patterns that are seen as reflexes of IVD effects. If IVD differences can be phonologized, then why is the corresponding sound change not attested more often? The question of underphonologization is an important one (Moreton 2008; 2010; Yu 2011), although a clear answer is generally hard to come by. Below we offer a few hypotheses.

To begin with, typologically speaking, languages generally do not contrast more than two degrees of vowel duration, despite notable exceptions (e.g., Estonian). Languages with three degrees of vowel height are very common; according to the World Atlas of Language Structures Online (Dryer & Haspelmath 2013), more than 80% of the languages surveyed (N = 564) have a vowel inventory of at least 5. The development of a three-way vowel length distinction based on vowel height would be against general typological preference for a binary vowel length distinction. Another variable worth considering is the fact that, out of the ninety-five participants examined in this study, only forty-nine exhibited a significant HEIGHT effect and only twenty-nine of which exhibited significant HEIGHT and TRIAL effects simultaneously. While we argue above that a constant duration difference strategy to maintain IVD across speaking rates is evidenced for individual-specific phonologization of IVD, the duration difference might nonetheless be small enough to prevent IVD from developing into a categorical height-dependent vowel length contrast. The "just-noticeable-difference" (JND) for duration discrimination roughly equals the square root of the interval (Stevens 2000: 228). Thus, for an average high vowel duration of about 200 ms, a duration difference of 12 ms between high and mid vowels observed in this study might only be barely enough to signal a JND in a discrimination task. To the extent that sound categories are learned distributionally (e.g., Maye et al. 2002; Clayards 2008), the close distribution of duration values between vowels of different heights might have prevented listeners from relying on the duration cue as a primary cue for the vowel height distinctions (Clayards et al. 2008). To be sure, one subject (i.e., Subject 51) did maintain IVD with a constant duration ratio strategy, suggesting that this speaker's IVD might be more noticeable at faster speaking rates. Nonetheless, the fact that only one out of ninety-five subjects maintains IVD using a constant duration ratio strategy might explain why IVD does not propagate across speakers so readily.

Our findings thus suggest that, despite the fact that HEIGHT is a significant predictor of vowel duration at the group-level analysis, IVD might not be as pervasive a pattern within the community of Cantonese speakers as the group results would suggest. Thus, despite the extensive reporting of IVD cross-linguistically, it remains a question whether IVD is widespread within each language community, especially since earlier studies have generally relied on a much smaller sample of speakers than this study. Finally, it is worth bearing in mind that reports of sound change generally rely on transcription systems that are ill-equipped to account for subtle duration differences between sounds. Take, for example, the fact that duration is a significant secondary cue for the tense and lax

vocalic contrast in English (e.g., Gordon et al. 1993; Escudero & Boersma 2004), yet phonological descriptions of English regularly ignore this difference in the representations of tense and lax vowels. This practice of ignoring subtle duration cues reflects a bias that assumes, often without argument, such duration differences as reflexes of non-linguistic differences attributable to physiological contingencies. It is precisely this bias that this study is meant to address. As revealed in their study of perceptual cue weights for the tense/lax distinction in English, Escudero & Boersma (2004) found that some native speakers of Southern British English actually rely on the duration cue as the primary cue over spectral ones. Given that most languages around the world have not been subjected to rigorous experimental examination to investigate the controlled vs. automatic nature of phonetic variation, the dearth of reported cases of IVD-related sound change might be more apparent than real.

4 General discussion

To summarize, this paper advances an approach that reconceptualizes (or perhaps more aptly, spotlights) phonologization as a matter of individual differences in acquiring controlled phonetic knowledge. From this perspective, phonologization does not involve individuals/groups of individuals transitioning between stages nor does it rely on the accumulation of occasional deviant pronunciation variants to account for the stabilization of sound change. Rather, phonologization is reflected in the difference of linguistic knowledge across speakers.

The idea that the synchronic pool of phonetic variations might contain both variants coming from universal phonetic contingencies as well as language-specific (indeed, speaker-specific) phonetic variation is not new. As noted above, in his H & H theory of sound change (Lindblom et al. 1995), Lindblom and colleagues posited that listeners acquire pronunciation variants that result from functional-communicative considerations. Such phonetic variants are assumed to be derived from phonetic factors related to hyper- and hypo-articulation. Blevins (2004) was more explicit about the role of individual variability in contributing to the plethora of variants in pronunciation norms. As alluded to earlier, Blevins (2004) posited three mechanisms of change: CHANGE, CHANCE, and CHOICE. Of interest here is the last mechanism, CHOICE. It incorporates Lindblom's idea of pronunciation variants as a source of change. For example, an assimilatory change such as *np > mp could be an instance of CHOICE where the phonetic realization of *np includes both [np] and [mp]. The survival and propagation of a particular variant may stem from considerations beyond phonetic concerns, such as the token frequency of a variant. It is in relation to the CHOICE mechanism that Blevins addressed the question regarding the origins of phonetic variants. While she acknowledged that variants might come from universal and language-specific phonetics, she assumed no clear distinction between them at the earliest stage of sound change (pages 39-40). Both give rise to a range of surface forms in natural speech production.

The perspective advocated in this study differs from Blevins' on two fronts. To begin with, it is important to maintain a distinction between automatic phonetic variation from the controlled ones, not only because they are theoretically distinct, are also demonstrably different both qualitatively and quantitatively, as Solé and others have shown (see, for example, Cohn 1990; Solé 1992; 2007; Solé & Ohala 2010). Second, individuals who have internalized controlled phonetic variations provide an essential stabilizing force, an anchor as it were, in introducing a potentially novel variant to the speech community. Pronunciation variations that are due to biomechanical forces or misperceived pronun-

ciations are likely to exhibit a much more variable distribution and might not allow the kind of stable social-indexical meaning association to take hold.[7] Finally, it is important to identify the individuals who have acquired controlled variants from those who have not since this distinction can offer insights into such individuals' roles as potential linguistic innovators in the community. For example, identifying such individuals can help isolate potential cognitive differences between individuals that have consequence for phonologization, particularly in light of recent studies that have highlighted the effect individual differences in cognitive processing styles have on speech and language (Stewart & Ota 2008; Ladd et al. 2013; Turnbull 2015; Jun & Bishop 2015; Yu 2016). Some processing style differences have been found to correlate with personality and social behavioral differences (Yu 2013). These social traits could inform how such individuals interact with their social network.

As already alluded to above, the recognition of the importance of the individual in understanding sound change is not new. However, the interpretations regarding the significance of individual variation in phonetic variation may differ. The listener misperception model of sound change and the H & H theory of sound change, for example, both point to the listener-turn-speaker as the locus of change. Various studies have examined individual variability in cue weighting (e.g., Hazan & Rosen 1991; Escudero & Boersma 2004; Kim & Hazan 2010; Shultz et al. 2012; Clayards 2018b; a; Coetzee et al. 2018). Much work has also focused on individual variability in coarticulation and the ramifications for sound change (e.g., Mann & Repp 1980; Repp 1981; Beddor 2009; Yu 2010; Yu & Lee 2014; Beddor 2015; Beddor et al. 2013; 2018; Mielke et al. 2016; Stevens & Harrington 2016; Yu 2016; 2019). Until recently, however, most of the earlier work on individual variability in speech perception and production has not theorized the significance of such variability for sound change research. In the research where such a connection is explicitly made (e.g., Beddor 2009; Yu 2010; Yu & Lee 2014; Beddor et al. 2013; Beddor 2015; Pinget 2015; Beddor et al. 2018; Mielke et al. 2016; Stevens & Harrington 2016; Yu 2016; Coetzee et al. 2018; Kuang & Cui 2018; Yu 2019), the authors do not always agree on the significance of individual variation or its connection to phonologization. Harrington and colleagues have investigated various cases of sound changes in progress both at the population and the individual levels. Stevens & Harrington (2016), for example, investigated the case of /s/-retraction in clusters containing a rhotic (e.g., *string*) in Australian English. They observed that all speakers exhibit spectral center of gravity lowering in /str/ and listeners would report hearing /s/ when the sibilant from /str/ clusters were spliced out and pre-pended to /iːt/. Crucially, they found that the magnitude of the acoustic effect of the /str/ context differed across individuals, even though all speakers exhibited some degree of retraction. This state of affairs is reminiscent of the case of individual variability observed in vowel-dependent /s/ variation in Hong Kong Cantonese examined in Yu (2016). While some early descriptions suggest a categorical /s/ to [ʃ] shift before rounded vowels, Yu (2016) found a high degree of variability across individuals in terms of the degree and temporal dynamic of the vocalic influence on /s/ realization. More importantly, the range of shifted /s/ does not resemble typical /ʃ/ found in other languages; Cantonese does not have a phonemic contrast between /s/ and /ʃ/. Interestingly, unlike Yu (2016), who interpreted the vowel-dependent /s/ variation as a case of controlled phonetic variation that has been phonologized by different individuals differently and might be, if only partially, related to autistic-like traits and gender, Stevens & Harrington (2016) concluded that /s/-retraction is "a shared, gradient

---

[7] It should be noted that, admittedly, it remains an open question whether haphazardly encountered innovative variants can be associated with socio-indexical meaning that could affect change.

tendency in the production of /str/ clusters" (123). They maintained that Australian English shows only the pre-conditions for /s/-retraction but is not currently undergoing the change.

It should be noted that the approach to phonologization laid out in this study has been anticipated in some earlier studies. As mentioned in the introduction, Beddor (2009), in her investigation on how individuals produce and perceive coarticulatory information in vowel-nasal sequences in English, found evidence that individuals differ in how they weigh the importance of coarticulatory cues in their perceptual responses and concluded that individuals might have different perceptual grammars (see also Beddor 2015; Beddor et al. 2013; 2018). Without explicitly referring to it as such, the idea that individuals might set up different perceptual grammars amounts to the idea that individuals have phonologized the timing relationship of nasality with its segmental anchors differently. Another line of inquiry that appeals to individual variation is agent-based simulation approaches to sound change. Stevens et al. (2019), for example, took as the starting point of their agent-based simulation models (see also Kirby & Sonderegger 2015; Harrington & Schiel 2017; Harrington et al. 2018) that "speakers are phonetically idiosyncratic even if they are of the same community and spoken accent; and that occasionally diachronic change can emerge if the idiosyncrasies of multiple speakers happen to fall along the path of a shared phonetic asymmetry" (2). Thus, different computational agents are initialized with "pronunciation" targets that show varying degrees of contextual influence, be it /u/-fronting (Harrington & Schiel 2017) or /s/-retraction (Stevens et al. 2019). Degree of shifts is examined after the agents are allowed to interact with each other in specific ways. These models have in common with the present model the emphasis on the interaction between individual speakers who have differing (context-dependent) acoustic/articulatory targets. To be sure, agent-based simulation studies emphasize understanding the role of individual variability in the propagation of sound change in progress; here, we emphasize understanding the nature of the initial state. That is, the present program emphasizes a conceptual commitment to the idea that individual differences in phonetic variations are par for the course within any population, and that understanding the non-random nature of such individual-specific variability is key to understanding the origins/phonologization of sound change.

The focus on the individual is also partly anticipated in studies that advocate modeling speech perception and production as governed, if only partly, by phonological considerations. Couched within Optimality-theoretic terms, for example, some scholars have argued that learners can posit phonetically-based constraints whose ranking/weighting influences the way phonological categories are setup (e.g., Escudero & Boersma 2004; Boersma 2006; Hamann 2009; Flemming 2010; Ramsammy 2015; 2018). While these models have in common the idea that many, if not all, phonetic processes are controlled, and may be modeled using theoretical tools similar to those developed for categorical phonological processes, it remains unclear the extent of individual variation and why such variability exists. To the extent that this is discussed, previous authors have generally attributed this variability, as well as the impetus for a phonetic precursor to transition into a categorical process, to differences in listener exposure. Ramsammy (2018), in his discussion of the life cycle of spirantization, for example, suggests that a change in status of a gradient phonetic process may come about as a consequence of children being exposed to adult speech models that include a sufficiently high number of fricated intervocalic tokens.[8] A later generation of speakers "may internalise a non-cognitively-

---

[8] Ramsammy made allowance that this change in status might extend beyond children who are acquiring the language to adults of the language as well, albeit subconsiously (see discussion on p.86 of that paper).

controlled stop frication pattern as a planned, stylistically dependent phonetic feature" (Ramsammy 2018: 87). As he summarizes, "a pattern of variation that is not under cognitive control in the speech generation n may phonologise into a cognitively-controlled pattern in the speech of generation $n + x$ by the pathway"(87), similar to the one laid out by Hyman above. Understanding the mystery behind how speakers, later generation or not, "may internalize a non-cognitively-controlled" phonetic pattern as planned is precisely what motivates the present approach to phonologization. Specifically, we question the assumption that there exists a generation of speakers where spirantization has not been phonologized. Why, for example, would the adult speech models include so many tokens of spirantization in the first place if spirantization has not actually been phonologized by some members within the community already? More research is needed to ascertain whether the assumption of listener exposure is necessary and sufficient to explain the existence of individual variability.

Understanding phonologization from an individual-difference perspective is admittedly a tall order as it requires the analyst to ascertain the internalized knowledge of individual speakers, rather than relying on group-level information to make generalizations regarding the community of speakers as a whole. In addressing the increased attention to phonetic variation and their relation to social factors, Labov (2006) once cautioned that "it does not follow automatically all such indexing should be described... Some further justification for the description of variation is required; otherwise there will be no stop to the enterprise and we will be plunged into an endless pursuit of detail" (508). We contend that it is amply justified to pay attention to individual variability, particularly in the context of understanding the nature of sound change. To be sure, it is not enough to merely describe the nature of the variation. As research in phonetics and phonology, and by extension, in sound change, becomes increasingly focused on individual variation, innovations will likely come in the form of more sensitive estimates of the individual's past linguistic experience (e.g., Lev-Ari & Peperkamp 2016), the biomechanical /articulatory specificity of individuals (e.g., Baker et al. 2011; Dediu & Moisik 2019; MacKenzie 2019), as well as speakers'/listeners' cognitive predispositions in terms of their processing styles and abilities (e.g., Stewart & Ota 2008; Lev-Ari & Peperkamp 2013; Niziolek & Guenther 2013; Yu 2013; Lev-Ari & Peperkamp 2014; Kong & Edwards 2016; Kapnoula 2016; MacKenzie 2019). Students of phonologization must pay attention to these individual-level factors and examine how they might play a role in affecting how speakers internalize phonetic variation as sound patterns.

In closing, the fact that systematic individual differences in phonetic variation exist, despite group-level patterning, suggests that sound change research cannot afford to focus on group-level differences alone. Furthermore, to the extent that differences between controlled vs. automatic phonetic processes exist within a single language community, it is important for linguists to identify them and understand how such differences come about. The eventual propagation of a controlled phonetic pattern to the rest of the population is contingent on the dynamics of the individuals within that community. As argued in Baker et al. (2011), sound change is predicted to be unlikely as it hinges on many contingencies to be present, including, but not limited to, the controlled phonetic variable being related to a socially relevant factor (Eckert 2019 and and references therein), and the individuals who exhibit the controlled phonetic patterns to be those others would like to emulate or imitate. Agent-based simulations of the type reviewed above are well-placed to tackle these complicated contingencies and interactions.

Abbreviations

F0 = Fundamental frequency, Hz = Hertz, IVD = Intrinsic vowel duration, JND = Just noticeable difference, L1 = First language, L2 = Second language, N = Nasal, V = Vowel, VOT = Voice Onset Time

Ethics and consent

Research involving human subjects were ethically conducted in accordance with the Declaration of Helsinki and the U.S. Code of Federal Regulations, and was approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago under IRB protocol number 12-1509.

Funding information

Acknowledgements

Competing interests

The author has no competing interests to declare.

References

Abramson, Arthur S. 1962. *The vowels and tones of standard Thai: Acoustical measurements and experiments*, vol. 20 (Indiana U. Research Center in Anthropology, Folklore, and Linguistics). Bloomington: Indiana University.

Äimä, Frans. 1918. *Phonetik und Lautlehre des Inarilappischen. I. Beobachtungsphonetik und deskriptive Lautlehre. II. Instrumentale Versuche und Messungen*, vol. 42–43 (Memoires de la Société Finno-Ougrienne). Helsinki: Suomalais-ugrilainen seura.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40(1). 177–189.

Baker, Adam, Diane Archangeli & Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23(3). 347–374.

Bates, Douglas, Martin Maechler & Ben Bolker. 2011. *lme4*. R package version 0.999375-38.

Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85(4). 785–832.

Beddor, Patrice Speeter. 2015. The relation between language users' perception and production repertoires. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings*

of the 18th international congress of phonetic sciences, 1–9. Glasgow, UK: International Phonetic Association. http://www.icphs2015.info/pdfs/Papers/ICPHS1041.pdf.

Beddor, Patrice Speeter, Andries Coetzee, Will Styler, Kevin McGowan & Julie Boland. 2018. The time course of individuals' perception of coarticulatory information is linked to their production: implications for sound change. *Language* 94. 1–38.

Beddor, Patrice Speeter, Kevin B. McGowan, Julie E. Boland, Andries W. Coetzee & Anthony Brasher. 2013. The time course of perception of coarticulation. *Journal of the Acoustic Society of America* 133(4). 2350–2366.

Bermúdez-Otero, Ricardo. 2007. Diachronic phonology. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 497–517. Cambridge: Cambridge University Press.

Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford, UK: Oxford University Press. DOI: 10.1093/oxfordhb/9780199232819.013.014.

Blevins, Juliette. 2004. *Evolutionary phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, Paul. 2006. Prototypicality judgments as inverted perception. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewky & Ralf Vogel (eds.), *Gradience in grammar*, 167–184. Oxford: Oxofrod University Press.

Chodroff, Eleanor, Alessandra Golden & Colin Wilson. 2019. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America* 145(1). EL109–115.

Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47. Https://doi.org/10.1016/j.wocn.2017.01.001.

Chodroff, Eleanor & Colin Wilson. 2018. Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard* .

Clayards, Meghan. 2008. *The ideal listener: making optimal use of acoustic-phonetic cues for word recognition*. Rochester, NY: University of Rochester dissertation.

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 EL* 144. Https://doi.org/10.1121/1.5052025.

Clayards, Meghan. 2018b. Individual talker and token variability in multiple cues to stop voicing. *Phonetica* 75. 1–23. Doi: 10.1159/000448809.

Clayards, Meghan, Michael K. Tanenhaus, Richard N. Aslin & Robert A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108. 804–809.

Coetzee, Andries W., Patrice Speeter Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in Afrikaans: differential cue weighting and sound change. *Journal fof Phonetics* 66. 185–216.

Cohn, Abigail. 1990. *Phonetic and phonological rules of nasalization*. Los Angeles, CA: UCLA dissertation.

Croft, William. 2000. *Explaining language change : An evolutionary approach*. Londres: Longman.

Dediu, Dan, Rick Janssen & Scott R. Moisik. 2017. Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. *Language & Communication* 54. 9–20. http://dx.doi.org/10.1016/j.langcom.2016.10.002.

Dediu, Dan & Scott R. Moisik. 2019. Pushes and pulls from below: Anatomical variation, articulation and sound change. *Glossa: a journal of general linguistics* 4(1). 7. DOI: https://doi.org/10.5334/gjgl.646.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/.

Eckert, Penelope. 2019. The individual in the semiotic landscape. *Glossa: a journal of general linguistics* 4(1). 1. DOI: https://doi.org/10.5334/gjgl.640.

Escudero, Paola & Paul Boersma. 2004. Bridging the gap between l2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551–585.

Fischer-Jorgensen, Eli. 1940. Objektive und subjektive Lautdauer deutscher Vokale. *Zeitschrift für die gesamte Phonetik* 4. 1–20.

Fischer-Jorgensen, Eli. 1964. Sound duration and place of articulation. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* 17. 175–207.

Flemming, Edward. 2010. Modeling listeners: Comments on Pluymaekers et al. and Scarborough. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.), *Papers in laboratory phonology 10*, 587–606. Berlin, Germany: Mouton de Gruyter.

Francis, Alexander L., Natalya Kaganovich & Courtney Discoll-Huber. 2008. Cue-specific effects of categorization training on therelative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America* 124(2). 1234–1251.

Gandour, Jack. 1977. On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica* 34. 54–65.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology.* Cambridge (MA): Basil Blackwell.

Gordon, Peter C., Jennifer L. Eberhardt & Jay G. Rueckl. 1993. Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology* 25. 1–42.

Gussenhoven, Carlos. 2004. Perceived vowel duration. In H. Quené & V. van Heuven (eds.), *On speech and language: Studies for Sieb G. Nooteboom*, 65–71. Utrecht: LOT.

Hale, Mark. 2007. *Historical linguistics: Theory and method.* Wiley-Blackwell.

Hale, Mark, Madelyn Kissock & Charles Reiss. 2015. An i-language approach to phonologization and lexification. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 337–358. Oxford, UK: Oxford University Press.

Hamann, Silke. 2009. The learner of a percepton grammar as a source of sound change. In Paul Boersma & Silke Hamann (eds.), *Phonology in perception* (Phonology and Phonetics 15), 111–149. Berlin: Mouton de Gruyter.

Harrington, Jonathan. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34. 439–457.

Harrington, Jonathan, Felicitas Kleber, Ulrich Reubold, Florian Schiel & Mary Stevens. 2018. Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science* 1–21.

Harrington, Jonathan, Sallyanne Palethorpe & Catherine I. Watson. 2000. Does the Queen speaker the Queen's English? *Nature* 408. 927–928.

Harrington, Jonathan & Florian Schiel. 2017. /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language* 93(2). 414–445.

Hazan, Valerie & Stuart Rosen. 1991. Individual variability in the perception of cues of place contrasts in initial stops. *Perception & Psychophysics* 49(2). 187–200.

Heffner, Roe-Merrill Secrist. 1937. Notes on the length of vowels. *American Speech* 12(2). 128–134.

House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25. 268–277.

Hyman, Larry. 1976. Phonologization. In Alphonse Juilland (ed.), *Linguistic studies presented to Joseph H. Greenberg,* 407–418. Saratoga: Anma Libri.

Idemaru, Kaori, Lori L. Holt & Howard Seltman. 2012. Individual differences in cue weight are stable across time: the case of Japanese stop lengths. *Journal of the Acoustical Society of America* 132(6). 3950–3964.

Johnson, Paul C. D. 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* 5(9). 944–946. DOI: 10.1111/2041-210X.12225.

Joseph, Brian D. 2011. Historical linguistics and sociolinguistics – strange bedfellows or natural friends? In Steffan Davies & Wim Vandenbussche. Nils Langer (ed.), *Language and history, linguistics and historiography,* 67–88. Peter Lang.

Jun, Sun-Ah & Jason Bishop. 2015. Priming iimplicit prosody: prosodic boundaries and individual differences. *Language and speech* 1–15. DOI: 10.1177/0023830914563368.

Kapnoula, Efthymia C., Matthew B. Winn, Eun Jong Kong, Jan Edwards & 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.

Kapnoula, Efthymia E. 2016. *Individual differences in speech perception: sources, functions, and consequences of phoneme categorization gradiency.* Iowa City, IA: University of Iowa dissertation.

Kim, Yoon Hyun & Valerie Hazan. 2010. Individual variability in the perceptual learning of l2 speech sounds and its cognitive correlates. In *Proceedings of the 6th international symposium on the acquisition of second language speech, new sounds 2010,* 251–256.

Kirby, James. 2010. *Cue selection and category restructuring in sound change.* Chicago, IL: University of Chicago dissertation.

Kirby, James & Morgan Sonderegger. 2015. Bias and population dynamics in the actuation of sound change. *arXiV* 1507. 04420.

Kong, Eun Jong & Jan Edwards. 2016. Individual differences in categorical perception of speech: cue weighting and executive function. *Journal of Phonetics* 59(1). 40–57.

Kong, Eun Jong & Hyunjung Lee. 2017. Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech* 1–25. Https://doi.org/10.1177/0023830917729840.

Kong, Qing Ming. 1987. Influence of tones upon vowel duration in Cantonese. *Language and Speech* 30(4). 387–399.

Kuang, Jianjing & Aletheia Cui. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71. 194–214.

Labov, William. 2006. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics* 34. 500–515.

Ladd, D. Robert, Rory Turnbull, Charlotte Browne, Catherine Caldwell-Harris, Lesya Ganushchak, Kate Swoboda, Verity Woodfield & Dan Dediu. 2013. Patterns of individual differences in the perception of missing-fundamental tones. *Journal of Experimental Psychology: Human Perception and Performance* 1–12. Doi: 10.1037/a0031261.

Lev-Ari, Shiri & Sharon Peperkamp. 2013. Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics* 41. 320–331.

Lev-Ari, Shri & Sharon Peperkamp. 2014. The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics* 47. 36–46.

Lev-Ari, Shri & Sharon Peperkamp. 2016. How the demographic makeup of our community influences speech perception. *Journal of the Acoustical Society of America* 139(6). 3076–3087.

Lindblom, Björn. 1967. Vowel duration and a model of lip-mandible coordination. *STL-QPRS* 4. 1–29.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling*, 403–439. The Netherlands: Kluwer.

Lindblom, Björn, Susan Guion, Susan Hura, Seung-Jae Moon & Raquel Willerman. 1995. Is sound change adaptive? *Rivista di Linguistica* 7. 5–36.

Lisker, Leigh. 1974. On "explaining" vowel duration variation. *Haskins Laboratories Technical Report* SR-37/38.

Maack, Adalbert. 1949. Die spezifische Lautdauer deutscher Sonaten. *Zeitschrift fur Phonetik* 3. 190–232.

MacKenzie, Laurel. 2019. Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa: a journal of general linguistic* 4(1). 28. DOI: https://doi.org/10.5334/gjgl.622.

Mann, Virginia A. & Bruno H. Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28. 213—228.

Maye, Jessica, Janet F. Werker & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3). B101–B111.

Mielke, Jeff, Adam Baker & Diana Archangeli. 2016. Individual-level contact limits phonological complexity: evidence from bunched and retroflex /ɹ/. *Language* 92(1). 101–140.

Milroy, James & Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21(2). 339–384.

Mirman, Daniel. 2014. *Growth curve analysis and visualization using r*. Boca Raton, Florida: Chapman and Hall / CRC.

Moreton, Eiliot. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127.

Moreton, Elliott. 2010. Underphonologization and modularity bias. In Steven Parker (ed.), *Phonological argumentation*, 79–101. London: Equinox Publishing.

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2). 133–142. DOI: 10.1111/j.2041-210x.2012.00261.x.

Navarro Tomás, Tomás. 1916. Cantidad de las vocales acentuadas. *Revista de Filología Española* 3. 387–408.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2). 132–142.

Niziolek, Caroline A. & Frank H. Guenther. 2013. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience* 33(41). 16110–16116.

Nooteboom, Sieb G. & I. Slis. 1970. A note on the degree of opening and the duration of vowels in normal and "pipe" speech. *IPO Annual Report* 5. 55–58.

Ohala, John J. 1981. The listener as a source of sound change. In & Mary F. Miller Carrie S. Masek, Roberta A. Hendrick (ed.), *Papers from the parasession on language and behavior*, 178–203. Chicago, IL: Chicago Linguistics Society.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer-Verlag.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv E. Breivik & Ernst H. Jahr (eds.), *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.

Ohala, John J. 1992. What's cognitive, what's not, in sound change. In Günter Kellerman & Michael D. Morrissey (eds.), *Diachrony within synchrony: Language history and cognition*, 309–355. Frankfurt: Peter Lang Verlag. Reprinted in Lingua e Stile (1992) 27: 321–362.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman Academic.

Ohala, John J. 1994. The frequency code underlies the sound symbolic use of voice pitch. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 325–347. Cambridge: Cambridge University Press.

Ohala, John J. 1995. Experimental phonology. In John A. Goldsmith (ed.), *A handbook of phonological theory*, 713–722. Oxford: Blackwell.

Ou, Jinghua & Alan C. L. Yu. 2019. Brainstem encoding of voice onset time: preliminary findings. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th international congress of phonetic sciences, melbourne, australia 2019*, 2114–2118. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Ou, Jinghua & Alan C. L. Yu. 2020. Neural correlates of individual differences in speech categorization: Evidence from subcortical, cortical, and behavioral measures.

Ou, Jinghua, Alan C. L. Yu & Ming Xiang. In press. Individual differences in categorization gradience as predicted by online processing of phonetic cues during spoken word recognition: Evidence from eye movements. *Cognitive Science*.

Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24. 175–184. (Data downloadable from http://www-2.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/speech/database/pb).

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustic Society of America* 32. 693–703.

Pinget, Anne-France. 2015. *The actuation of sound change*. Utrecht, the Netherlands: LOT.

Ramsammy, Michael. 2015. The life cycle of phonological processes: accounting for dialectal microtypologies. *Language and Linguistics and Language Compass* 9(1). 33–54.

Ramsammy, Michael. 2018. The phonology-phonetics interface in constraint-based grammar: gradience, variability, and phonological change. In S.J. Hannahs & Anna R. K. Bosch (eds.), *The Routledge handbook of phonological theory*, Abingdon: Routledge.

Repp, Bruno H. 1981. Two strategies in fricative discrimination. *Perception & Psychophysics* 30(3). 217–227.

Sachs, Jacqueline, Phillip Lieberman & Donna Erickson. 1973. Anatomical and cultural determinants of male and female speech. In Roger W. Shuy & Ralph W. Fasold (eds.), *Language attitudes: Current trends and prospects*, 74–84. Washington, DC: Georgetown University Press.

Sankoff, Gillian & Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3). 560–588.

Scharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5. 26–30.

Schertz, Jessamyn, Taehong Cho, Andrew Lotto & Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52. 183–204.

Shultz, Amanda A., Alexander L. Francis & Fernando Llanos. 2012. Differential cue weighting in perception and production of consonanat voicing. *Journal of Acoustical Society of America Express Letter* 132(2). 95–101. Http://dx.doi.org/10.1121/1.4736711.

Smith, Bridget J., Jeff Mielke, Lyra Magloughlin & Eric Wilbanks. 2019. Sound change and coarticulatory variability involving English /ɹ/. *Glossa: a journal of general linguistics* 4(1). 63. DOI: http://doi.org/10.5334/gjgl.650.

Smith, Bruce L. 1987. Effects of bite block speech on intrinsic segment duration. *Phonetica* 44(2). 65–75.

Solé, Maria-Josep. 1992. Phonetic and phonological processes: The case of nasalization. *Language and Speech* 35(1-2). 29–43.

Solé, Maria-Josep. 1995. Spatio-temporal patterns of velo-pharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38(1). 1–23.

Solé, Maria-Josep. 2007. Controlled and mechanical properties in speech: a review of the literature. In Maria-Josep Solé, Patrice Speeter Beddor & Manjuri Ohala (eds.), *Experimental approaches to phonology*, 302–321. Oxford: Oxford University Press.

Solé, Maria Josep & John J. Ohala. 2010. What is and what is not under the control of the speaker: intrinsic vowel duration. In Barbara Kühnert Cécile Fougero, Mariapaola Di'Imperio & Nathali Vallée (eds.), *Laboratory phonology 10*, Mouton de Gruyter.

Stevens, Kenneth N. 2000. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, Mary & Jonathan Harrington. 2014. The individual and the actuation of sound change. *Loquens* 1(1). e003. Doi: 10.3989/loquens.2014.003.

Stevens, Mary & Jonathan Harrington. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118–134.

Stevens, Mary, Jonathan Harrington & Florian Schiel. 2019. Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa: A Journal of General Linguistics* 4(1). 8. DOI: http://doi.org/10.5334/gjgl.620.

Stewart, Mary E. & Mitsuhiko Ota. 2008. Lexical effects on speech perception in individuals with "autistic" traits. *Cognition* 109. 157–162.

Stone, Adam. 2014. Vowel height and duration. Institute of Cognitive Science Annual Spring Conference, Carleton University.

Tauberer, Joshua & Keelan Evanini. 2009. Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. In *Proceedings of INTERSPEECH 2009 – 10th annual conference of the International Speech Communication Association, Brighton, UK, September 6–10, 2009*, 2211–2214. ISCA. Http://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_2211.pdf.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Toivonen, Ida, Lev Blumenfeld, Andrea Gormley, Leah Hoiting, John Logan, Nalini Ramlakhan & Adam Stone. 2015. Vowel height and duration. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, 64–71. Somerville, MA, USA: Cascadilla Proceedings Project.

Turnbull, Rory. 2015. Pattterns of individual differences in reduction: implications for listener-oriented theories. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. Glasgow, UK: The University of Glasgow. http://www.icphs2015.info/pdfs/Papers/ICPHS0106.pdf.

Vorperian, Houri K., Shubing Wang, E. Michael Schimek, Reid B. Durtschi, Ray D. Kent, Lindell R. Gentry & Moo K. Chung. 2011. Development sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language and Hearing Research* 54. 995–1010.

Wickham, Hadley. 2012. *plyr*. R package version 1.8.

Wilson, Colin & Eleanor Chodroff. 2017. Uniformity of inherent vowel duration across speakers of American English. Poster at the 174th Meeting of the Acoustical Society of America, New Orleans, LA.

Yu, Alan C. L. 2007. Understanding near mergers: The case of morphological tone in Cantonese. *Phonology* 24(1). 187–214.

Yu, Alan C. L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS One* 5(8). e11950. doi:10.1371/journal.pone.0011950.

Yu, Alan C. L. 2011. On measuring phonetic precursor robustness: A response to Moreton 2008. *Phonology* 28(3). 491–518. University of Chicago.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, 201–227. Oxford, UK: Oxford University Press.

Yu, Alan C. L. 2016. Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *Journal of Acoustical Society of America* 139(4). 1672–1690.

Yu, Alan C. L. 2019. On the nature of the perception-production link: Individual variability in english sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1). 2. DOI: http://doi.org/10.5334/labphon.97.

Yu, Alan C. L. & Hyunjung Lee. 2014. The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *Journal of the Acoustical Society of America* 136(1). 382–388.

Yu, Alan C. L., Hyunjung Lee & Jackson Lee. 2014. Variability in perceived duration: pitch dynamics and vowel quality. In Carlos Gussenhoven, Yiya Chen & Dan Dediu (eds.), *The 4th international symposium on tonal aspects of languages*, 41–44. Nijmegen, The Netherlands: ISCA Archive. http://www.isca-speech.org/archive/tal_2014/tl14_041.html.

Zellou, Georgia. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61(1). 13–29.