Special Issue: Marking 50 Years of Research on Voice Onset Time, eds. Cho, Docherty & Whalen

Voice onset time and beyond: Exploring laryngeal contrast in 19 languages

Taehong Cho [a,*], D.H. Whalen [b,c,d], Gerard Docherty [e]

[a] Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS), Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdon-gu, Seoul 04763, Republic of Korea
[b] Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, United States
[c] Program in Speech-Language-Hearing Sciences, Graduate Center, City University of New York, 365 5th Ave., New York, NY 10016, United States
[d] Department of Linguistics, Yale University, PO Box 208366, New Haven, CT 06520-8366, United States
[e] Arts, Education and Law Group, Griffith University, 176 Messines Ridge Road, Mt Gravatt Campus Q 4122, Brisbane, Australia

A R T I C L E   I N F O

A B S T R A C T

In this special collection entitled *Marking 50 Years of Research on Voice Onset Time and the Voicing Contrast in the World's Languages*, we have compiled eleven studies investigating the voicing contrast in 19 languages. The collection provides extensive data obtained from 270 speakers across those languages, examining VOT and other acoustic, aerodynamic and articulatory measures. The languages studied may be divided into four groups: 'aspirating' languages with a two-way contrast (English, three varieties of German); 'true voicing' languages with a two-way contrast (Russian, Turkish, Brazilian Portuguese, two Iranian languages Pashto and Wakhi); languages with a three-way contrast (Thai, Vietnamese, Khmer, Yerevan Armenia, three Indo-Aryan languages, Dawoodi, Punjabi and Shina, and Burushaki spoken in India); and Indo-Aryan languages with a more than three-way contrast (Jangli and Urdu with a four-way contrast, and Sindhi and Siraiki with a five-way contrast). We discuss the cross-linguistic data, focusing on how much VOT alone tells us about the voicing contrast in these languages, and what other phonetic dimensions (such as consonant-induced F0 and voice quality) are needed for a complete understanding of laryngeal contrast in these languages. Implications for various issues emerge: universal phonetic feature systems, effects of language contact on linguistic levelling, and the relation between laryngeal contrast and supralaryngeal articulation. The cross-linguistic VOT data also lead us to discuss how the distribution of VOT as measured acoustically may allow us to infer the underlying articulation and how it might be approached in gestural phonologies. The discussion on these multiple issues sparks new questions to be resolved, and provide indications of where the field may be best directed in exploring laryngeal contrast in voicing in the world's languages.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

It has been just over a half century since Lisker and Abramson (1964) proposed an acoustic measure of Voice Onset Time (VOT) as a unitary and eminently tractable basis on which to characterize the voicing categories of stops across languages which had often been distinguished by seemingly independent phonetic features of voicing, aspiration and "force of articulation." Based on observations of voicing patterns in eleven languages, Lisker & Abramson made a key assumption that a fairly complicated acoustic output in association with

different voicing categories within and across languages arises as a predictable consequence of varying the area of the glottis. The underlying laryngeal setting was proposed to be effectively captured by VOT defined by the "relative timing of events at the glottis and at the place of oral occlusion." There were later expansions to the application of VOT to include intervocalic stops (Abramson, 1977) and affricates (Abramson, 1995). Since then, this innovative measure has been adopted by virtually every experimental phonetic study that has investigated acoustic characteristics of stop consonants, thereby greatly advancing our understanding of voicing properties of stop consonants and their typology in the world's languages.

In a recent Technical Note submission to *Journal of Phonetics* (Abramson & Whalen, 2017, now also included as part of this special collection), Arthur Abramson and D. H. Whalen have provided a retrospective commentary entitled "Voice Onset Time (VOT) at 50: Theoretical and practical issues in

* Corresponding author at: Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS), Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Republic of Korea. Fax: +82 2 2220 0741.
E-mail addresses: tcho@hanyang.ac.kr (T. Cho), whalen@haskins.yale.edu (D.H. Whalen), gerry.docherty@griffith.edu.au (G. Docherty).

measuring voicing distinctions." It bore largely on procedural aspects of the application of VOT, its limitations, and ways to expand the notion of VOT to a wider range of different phonological contexts. Inspired by this retrospective reflection on VOT, we have commissioned a special collection of themed papers in order to mark the occasion of 50 plus years of VOT under the title "Marking 50 Years of Research on Voice Onset Time and the Voicing Contrast in the World's Languages." This special collection devotes itself to exploring the phonetic properties of voicing contrasts with a view to providing a contemporary lens on various aspects of consonantal voicing contrast within and across the world's languages from both theoretical and methodological perspectives, and relevant points of debate that have endured alongside or as an alternative to VOT.

1.1. Languages covered: 19 languages with 270 speakers

In this special collection, an impressive array of languages was studied by 19 authors over 11 papers (10 new submissions plus Abramson & Whalen, 2017). The 10 new submissions covered production data on the voicing contrast obtained from 270 speakers across 19 languages. Among the 19 languages studied are two languages whose voicing contrast has been well-documented in the literature: English appearing in two of the contributions (Ahn, 2018a with 8 speakers; Kim, Kim & Cho, 2018 with 11 speakers), and German, with three varieties appearing across two contributions (Swiss German: Ladd & Schmid, 2018 with 20 speakers; Bavarian and Saxon varieties of German: Kleber 2018 with 21 and 20 speakers, respectively). Other languages whose voicing contrast has not been fully understood despite the substantial number of their speakers include Brazilian Portuguese (Ahn, 2018a with 8 speakers); Thai (Kirby, 2018 with 12 speakers); Turkish (Ünal-Logacev, Fuchs & Lancia, 2018 with 6 speakers); and Russian (Kharlamov, 2018 with 60 speakers). Languages that have received even less attention but are covered in this special collection include Lebanese Arabic (Al-Tamimi & Khattab, 2018 with 20 speakers), Vietnamese and Khmer (Kirby, 2018 with 14 speakers each); Yerevan (Eastern) Armenian (Seyfarth & Garellek, 2018 with 8 speakers), and 10 languages (two Iranian, seven Indo-Aryan languages and one isolated one) spoken in India with 48 speakers in total (Hussain, 2018).

2. VOT as a first estimate of voicing contrast

Some of the fundamental questions that this special collection covers concern (1) how different patterns of phonetic realization of voicing that may occur in different segmental, phonological and prosodic contexts could be adequately described by employing the basic notion of VOT and its extension (e.g., Abramson & Whalen, 2017; Davidson, 2016); and (2) to what extent VOT alone would suffice or whether other phonetic parameters would be necessary to adequately capture the voicing contrast of consonants within and across languages.

As we have stated above, it goes without saying that VOT does not capture every acoustic aspect of voicing distinctions in stops (and other consonants), but the current collection of studies provide ample evidence that VOT serves as a useful first estimate of a language's use of laryngeal distinctions in voicing. All 19 languages (21 varieties with three German dialects) studied in this special collection utilize phonetic voicing as reflected in VOT differences in order to mark in one way or another phonological contrast among stops, although they differ typologically in terms of how many laryngeal distinctions in voicing are employed in the phonological system.

2.1. Languages with two-way contrast

Among the 19 languages studied, seven languages (English, German, Russian, Turkish, Brazilian Portuguese, Pashto and Wakhi) employ a two-way distinction for laryngeal contrast between stops, which may be phonologically classified as [voiced] *vs.* [voiceless]. The two-way system, however, is further divided into so-called 'true voicing' and 'aspirating' languages, depending on whether the [voiced] stops are produced with voicing lead before the release (prevoicing) or phonation during closure (in the case of true-voicing languages such as Russian, Turkish, Portuguese, Pashto and Wakhi) or the [voiceless] stops are produced with aspiration (in the case of aspirating languages such as English and German with three varieties, Swiss, Bavarian and Saxon). VOT data measured for all these languages (except for Swiss German) indicate that the two-way contrast can be adequately captured by polarization of voicing contrast along the VOT continuum. On the one hand, the true voicing languages show a substantial negative VOT or voicing lead (i.e., phonation (voicing) during closure) for the phonologically [voiced] stop category, whereas the phonologically [voiceless] stops in these languages are produced with a short-lag (positive) VOT. On the other hand, the aspirating languages such as English and Bavarian/Saxon German show the use of VOT primarily in the positive dimension—i.e., the phonologically [voiced] stops are by and large phonetically voiceless with a short-lag (positive) VOT especially in utterance-initial positions (though partially or fully voiced variants often occur in other, prosodically weak, positions in connected speech, cf. Davidson, 2018) while their voiceless counterparts are produced with a long-lag (positive) VOT which includes a measurable period of aspiration.

2.2. Languages with three-way contrast

Another eight languages employ a three-way stop contrast: Thai, Vietnamese, Khmer, Yerevan Armenian, Dawoodi, Punjabi, Shina and Burushaki. These languages, despite belonging to different language families, all disperse the three-way contrastive stops along the VOT continuum in a comparable way, corresponding to the three phonetic categories of {voiced}, {voiceless unaspirated} and {voiceless aspirated}. (Note that, following Keating, 1984, curly brackets '{ }' are used here to refer to phonetic features or categories as opposed to phonological features [+/−voice].) The voiced stops are produced with substantial phonation during closure as reflected in long negative VOTs (just as in the true voicing languages), and the voiceless unaspirated/aspirated stops are produced with a short-lag and a long-lag (positive) VOT, respectively. This means the whole range of the VOT continuum is employed quite exhaustively by these languages, showing a

sufficient dispersion along the VOT dimension (see Cho & Ladefoged, 1999, for related discussion).

2.3. Languages with more than a three-way contrast

The remaining four languages, all spoken in India, employ either a four-way stop contrast (Jangli and Urdu) or a five-way contrast (Sindhi and Siraiki). The four-way contrastive stops in Jangli and Urdu may be represented, for example in the case of the labial series, with /b bʱ p pʰ/ and the five-way contrastive stops with /b bʱ ɓ p pʰ/ (with an implosive (/ɓ/) added). These languages too employ VOT for signaling up to a three-way contrast (e.g., /b/-/p/-/pʰ/), whose VOT distribution is strikingly similar to that employed by languages with a three-way contrast. But VOT does not go beyond distinguishing more than three categories. When an additional voiced aspirated series (e.g., /bʱ/) is added in both the four-way and the five-way contrast systems, VOT of the voiced aspirated category overlaps substantially with that of voiced unaspirated stops (e.g., /b/) both distributed largely in the negative VOT dimension, but the voiced aspirated series in one of these languages (in Jangli) shows an even wider distribution of VOT, often extending to the positive territory of VOT. Finally, when the voiced implosives (e.g., /ɓ/) are added in Sindhi and Siraiki, the three voiced categories (e.g., /b bʱ ɓ/) all employ the negative VOT dimension with no clear VOT distinction among them. As Hussain noted, these constitute cases which require other phonetic parameters to understand how the stop categories with overlapping VOT distribution are further distinguished (see below for more discussion on this issue). The limits on the use of VOT in such cases were, as noted, already discussed in Lisker and Abramson (1964).

3. Universals and variation in VOT: evidence from 19 languages

3.1. Universal feature systems reflected in VOT

In accounting for phonological versus phonetic aspects of the voicing contrast, Keating (1984, 1990a) suggested that while the phonological feature (such as [± voice]) may be required to explain the phonological contrast between voiced and voiceless stops within each language (see also Kingston & Diehl, 1994), the phonetic features such as {voiced}, {vl. unasp.}, {vl. asp.} are required to explain how stops are phonetically implemented between languages. (Note that the two phonetic features {vl. unasp.} and {vl. asp.} in Keating (1984) were later replaced by {−spread glottis} and {+spread glottis} in Keating (1990a), features which are grounded on laryngeal articulatory characteristics rather than the acoustic output.)

As exemplified by some of the languages discussed above, a majority of languages may be classified as showing either a binary or a three-way phonological voicing contrast which may be adequately captured by how the VOTs of the phonetic categories are distributed among the VOT continuum. The distribution of VOT especially in those languages with a two-way and a three-way voicing contrast may indeed be largely predicted by how each phonological category is mapped on to one of the three universally available phonetic features, based on which phonetic realization is implemented (e.g., Keating, 1984, 1985; cf. Cho & Ladefoged, 1999). The 'true voicing'

languages are then taken to employ {voiced} and {vl. unasp.} reflected respectively in a negative VOT and a short-lag (positive) VOT, whereas the 'aspirating' languages use {vl. unasp.} and {vl. asp.} which are reflected by a short-lag VOT and a long-lag VOT in the positive dimension. Similarly, the VOT distribution in languages with a three-way voicing contrast can be taken to be mapped on to all three phonetic features as evident in remarkably similar ranges of VOTs across languages. What is particularly noteworthy is that even for the languages with a four-way or a five-way contrast, VOT makes a clear three-way distinction among the three categories mapped on to the phonetic features {voiced}, {vl. unasp.} and {vl. asp.} again with comparable ranges of VOT for each category found in those languages with a two-way or a three-way voicing contrast. The emerging cross-linguistic similarities in the distribution of VOTs suggest that VOT is an important metric for understanding the language universals underlying laryngeal contrast in voicing in the worlds' languages, and that such cross-linguistic similarities can be accounted for by the universally available three phonetic features mapped on to similar ranges of VOT in the languages studied in this special collection.

3.2. VOT as a controllable metric and the phonetic grammar

Just as much as languages are similar in using VOT for voicing contrast, they may be dissimilar in exactly where in the VOT continuum each of the opposing categories is anchored to signal the phonological contrast. Based on VOT distributions in 18 languages, Cho and Ladefoged (1999) (and Ladefoged & Cho, 2001) proposed a so-called 'Articulatory VOT' as a controllable variable defined as the timing between the supralaryngeal release gesture and the laryngeal voicing gesture. (Note that the term 'Articulatory VOT' was first used by Ladefoged and Cho (2001), but Lisker and Abramson (1964) also defined VOT from an articulatory point of view, although it has often been described as an acoustic measure.) The Articulatory VOT is assumed to be fine-tuned by the phonetic grammar of the language, yielding cross-linguistic differences. Based on this assumption, Cho and Ladefoged (1999) suggested that while languages generally follow the distribution of VOT within an allowable window specified by phonetic features (in the spirit of Keating's (1990b) window model), the fine-phonetic detail observed across languages cannot be accounted for merely by general phonetic principles such as ease of articulation (or "low-cost" options in Docherty's (1992) term) and contrast maximization (Lindblom, 1986, 1990). The cross-language data instead revealed language arbitrariness in choosing a 'modal' VOT within a language, which brings about variation across languages. For example, mean VOT values for velar stops vary from around 28 ms to 80 ms in 11 languages even though these languages do not make more than one phonological contrast along the positive VOT dimension. The VOT distribution over a wider range not only makes it hard to determine where to draw a clear-cut line between phonetically unaspirated and aspirated stops across languages, it is also at odds with the principle of maximal articulatory ease. The "low-cost" option for languages with no phonological contrast between unaspirated and aspirated stops would be to use a single, simplest articulatory gesture for the voiceless sound (Docherty, 1992). This would predict

very similar VOT values for languages. In fact, as discussed above, the languages included in this special collection appear to show remarkably similar VOT ranges for each category. For example, those languages such as Lebanese Arabic, Swiss German, Pashto, Wakhi, and Brazilian Portuguese that make no VOT distinction in the positive VOT dimension (i.e., only voiceless unaspirated stops) have a range of VOT from 1.4 to 21 as shown in Fig. 1a. These languages thus show a short-lag VOT that could conceivably be mapped on to the phonetic feature {vl. unasp.} (or {−spread glottis}). It is also noticeable that Brazilian Portuguese and American English belong to a 'true voicing' *vs.* an 'aspirating language' group, respectively, but the VOTs (taken from the same study, Ahn 2018a) are remarkably similar, both mapped on {vl. unasp.} even though they are associated with different phonological features: [+voice] (in English) vs. [−voice] (in Portuguese).

But there is also some variation across languages. As seen in Fig. 1, while Lebanese Arabic produced the shortest VOTs of 1.5–2.9 ms leading to a position on the left end of the continuum, Turkish represents a rightward deviation from the general pattern with a mean VOT of 41 ms being placed at the right end of the continuum. Note that the short VOTs in Lebanese Arabic were in part due to the fact that VOTs were taken from tokens some of which had partial voicing during closure. The mean VOT in Turkish shown in the figure was taken from alveolars (whose VOTs are relatively shorter than those of velars) and they were produced in a sentence frame. So the relatively longer VOT could not be attributable to either a place effect or a speech rate effect. Note also that the VOT values for Brazilian Portuguese and American English in Fig. 1a are based on alveolar stops produced in isolation. Thus, even if we take into account possible mismatches in various other factors (e.g., speech rate, position), we can safely conclude that the VOT in Turkish deviates clearly from those in the other languages shown in Fig. 1a.

The language-specific setting is more evident in VOT distributions of the voiceless aspirated stops in languages which employ a two-way phonetic distinction in the positive VOT dimension (i.e., those which have both the voiceless unaspirated and aspirated distinction). As shown in Fig. 1b, mean VOTs of aspirated (denti) alveolar stops in these languages are distributed over a wide range from 57 ms to 97 ms. Among these languages, 8 languages were studied by Hussain (2018) with a similar method. These languages also show a similar variation in mean VOT from 57 ms to 91 ms. Furthermore, the distribution of VOT does not seem to be affected by how many phonological distinctions are made within a language. For example, four Indo-Aryan languages (Jangli, Urdu, Sindhi, Siraiki) which employ four- or five-way contrasts do not necessarily show a greater polarization, relative to the languages (e.g., Punjabi, Dawoodi, Khmer, Burushaski, Thai) which show a three-way contrast. Thus, the VOT distribution of stops shown in Fig. 1b implies that although the VOTs of stops in these languages may be mapped on to the {vl. asp.} phonetic category, it supports the view that languages choose a modal VOT value of their own within a window that is permissible by the phonetic feature {vl. aspirated} (or {+spread glottis}) in line with Cho and Ladefoged (1999) and Ladefoged and Cho (2001).

The voiced categories in 'true voicing' languages further illuminate cross-linguistic similarities and differences. The distribution of (negative) VOTs of (denti-)alveolar stops in 17 languages
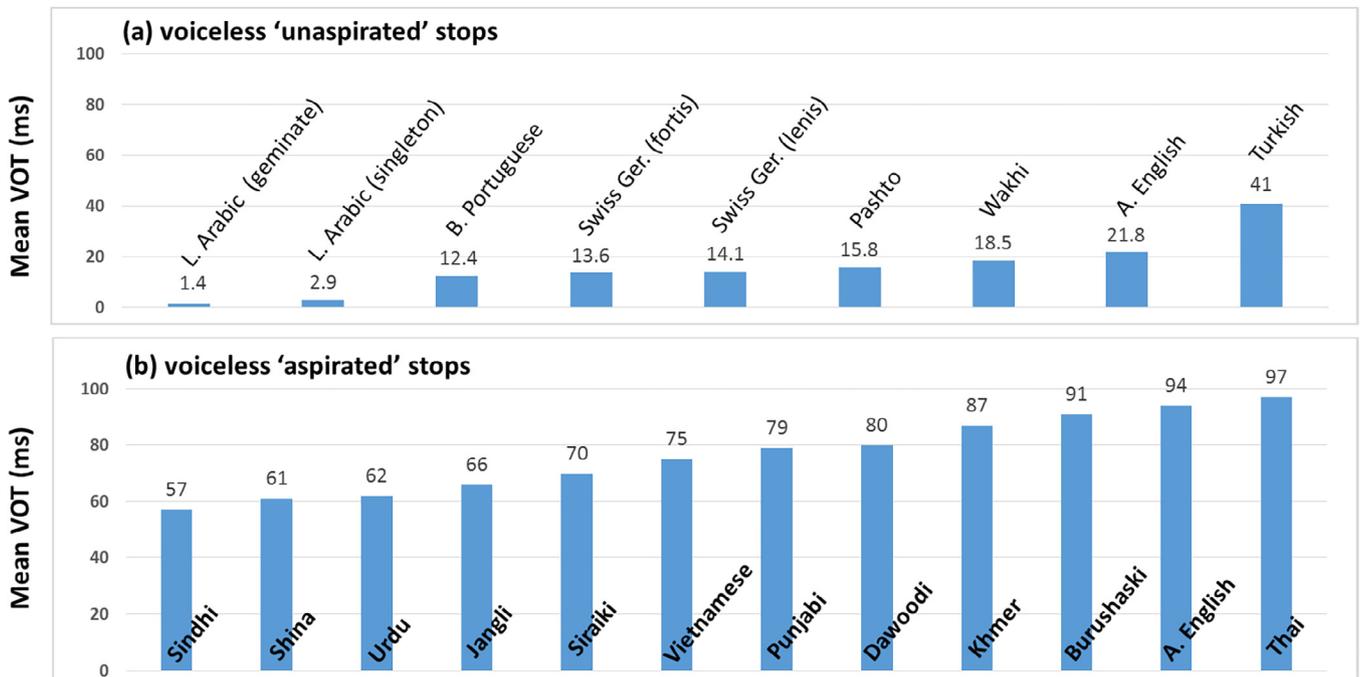

Fig. 1. Distribution of mean VOTs in languages studied in this special collection (see Section 1.1 for the references): (a) voiceless unaspirated category along the positive VOT dimension and (b) two voiceless categories along the positive VOT dimension. VOTs in three languages (Lebanese Arabic, Swiss German, Turkish) were obtained from words produced in a sentence frame, and the rest were from words in isolation. VOTs were taken from (denti-)alveolars, except for the ones in Swiss German whose values were pooled across different places of articulation. Note that tokens produced with some voicing during closure are excluded in Lebanese Arabic, Brazilian Portuguese and American English. (English data were taken from Ahn, 2018a, this collection). Note also that the mean VOT values for alveolar stops in Lebanese Arabic were obtained directly from the authors of Al-Tamimi and Khattab (2018, this collection) and in the original paper, they reported mean VOTs pooled across different places of articulation (8.7 ms for geminates and 5.3 ms for singletons).

Voiced (denti-alveolar) stops with prevoicing (phonation) during closure in 17 languages
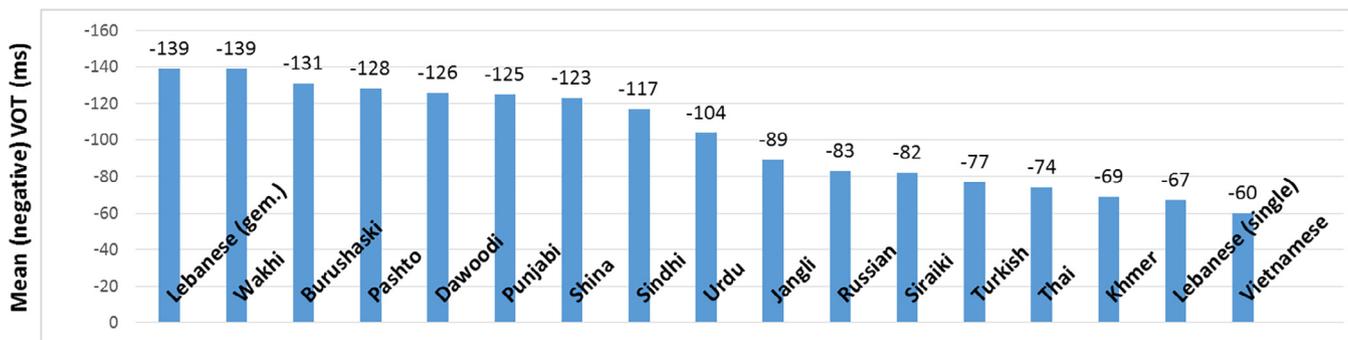

Fig. 2. Mean VOTs of phonetically voiced (denti-) alveolar stops in 17 languages studied in this special collection. Stops were produced in words in isolation in all languages but Yerevan Armenian for which they were produced in a frame sentence.

is given in Fig. 2. The negative VOTs of these languages are generally considered to be fully voiced (if we apply the 50% threshold of voicing during closure as suggested by Abramson and Whalen, 2017). These values may then be mapped on to the phonetic feature {voiced} which might require 'active' articulatory control for initiating and maintaining phonation during closure. The negative VOT data in Fig. 2 can thus be seen as showing some universal pattern of voicing realization within a permissible range of the phonetic feature {voiced} (i.e., more than 50% of the closure duration is phonetically voiced). But here again the actual temporal stretch of voicing varies across languages ranging from means of −139 ms (Wakhi) to −60 ms (Vietnamese) when alveolar consonants are considered. As was the case with the voiceless categories, the variation observed with the voiced stops is largely independent from how many distinctions are made in the phonological system of a given language. For example, Siraiki employs a five-way distinction with three categories to be contrastive along the negative VOT dimension. But the voiced (unaspirated) stop in this language is produced with a relatively short negative VOT, whereas Wakhi with a three-way distinction (with only one in the negative dimension) shows a relatively longer (negative) VOT, although these two languages belong to the same language family of Indo-Aryan. This builds on Cho & Ladefoged (1999) who discuss language arbitrariness of choosing VOT in the 'positive' dimension, and further implies that languages also choose a modal VOT in the 'negative' dimension but within a permissible window mapped on to the phonetic feature {voiced}.

Finally, a question that remains unanswered concerns how variation among speakers within a language can be accounted for in conjunction with such a language-specifically determined 'modal' VOT (assumed to be internalized in the phonetic grammar of each language). The cross-linguistic data reported in this special collection do not have a direct bearing on this issue, but Chodroff and Wilson (2017) provide VOT distribution data of stops /b d g p t k/ across multiple American English speakers that illuminate the nature of speaker variation. It was reported that while VOT values in absolute terms varied substantially among speakers, the speakers showed similar linear relations between voiced and voiceless stops (e.g., /b/ vs. /p/) and between different places of articulation (e.g., /p/ vs. /k/) both in isolation (from 24 speakers) and in connected speech (from more than 100 speakers). The pattern of VOT covariation was

interpreted as supporting a 'uniformity' constraint. It is assumed to restrict the speaker-specific realization of a phonetic property in a principled way, such that speaker variation is permissible in absolute terms insofar as structured relationship is maintained uniformly across speakers. When we consider such a uniformity constraint, we can further infer that a 'modal' VOT for a given language may be internalized in the phonetic grammar of the language in relational terms—i.e., in reference to structured speaker variation with some degree of relational invariance that arises with various factors that are known to influence phonetic realization of VOT such as place of articulation and strength of prosodic juncture (as reflected in domain-initial strengthening).

4. Beyond VOT

Thus far we have discussed the extent to which VOT alone can account for voicing contrasts that occur across languages which employ a two-way or a multi-way laryngeal contrast in voicing in their phonological systems. There are, however, cases in which voicing contrast cannot be fully captured by VOT alone, or cases in which the phonetic nature of voicing contrast can be further illuminated along phonetic dimensions other than VOT such as voice quality and F0. Many of the Indo-Aryan languages (Hussain, 2018) present such cases as they employ multiple distinctions made along the negative VOT dimension with voiced (unaspirated) stops in contrast with voiced aspirated stops and voiced implosives. Hussain (2018) suggests a number of possible phonetic correlates of the laryngeal distinction especially for the stops whose VOT values overlap substantially to the extent that no further distinction could be made along the negative VOT dimension. Ladd and Schmid (2018) pose another case in Swiss German in which VOT does not play a role in making the lenis-fortis distinction. Furthermore, except for Hussain (2018, Indo-Aryan and Iranian languages) and Kim et al (2018) (American English), all other studies have attempted to explore other possible phonetic dimensions (e.g., F0, voice quality, aerodynamics, supralaryngeal articulation) that may either replace or complement the voicing distinction signaled by VOT. In this section, we will focus on two such acoustic parameters, CF0 and voice quality that have also been examined quite extensively in the literature in relation to laryngeal contrast across languages.

4.1. CF0

CF0 (consonant-induced F0) refers to F0 of the following vowel that may vary systematically with the voicing contrast of the preceding consonant. The notion of CF0 originates from low-level (biomechanic) F0 perturbation that 'perturbs' the F0 at the onset of the following vowel especially by raising F0 after a voiceless stop. It is often considered to stem from tension of laryngeal muscles that are involved in production of voiceless consonants (e.g., Hanson 2009; Kohler, 1982; Kingston, 2007; Löfqvist, Baer, McGarr, & Story, 1989). (See Kingston and Diehl (1994) and Kingston (2007) for related discussion on the linguistic use of CF0). CF0 has also been the source of tonogenetic sound change in some languages (e.g., Seoul Korean and Afrikaans) in which the phonological voicing contrast, which used to be marked primarily by the VOT difference, is now marked primarily by the tonal difference (e.g., Bang, Sonderegger, Kang, Clayards, & Yoon, 2018; Coetzee, Beddor, Shedden, Styler, & Wissing, 2018; Kang, 2014; see Kingston, 2011 for a review).

Ladd and Schmid (2018, this collection) explore effects of phonological stop voicing contrast on CF0 in Swiss German and discuss how it would inform us about the use of CF0 in phonetic implementation of voicing contrast. A motivation for this study stems from the fact that the fortis and lenis stops in Swiss German are both produced as voiceless unaspirated stops with no discernible distinction in VOT, and the phonological contrast is generally considered to be phonetically signaled by the difference in closure duration. This poses a question for the fundamental assumption that a majority of voicing contrast in the world's languages may be characterized by identifying which (universally available) voicing-related features (e.g., [voice], [spread glottis], [tense]) are chosen by the language as discussed above (e.g., Keating, 1984, 1985, 1990a; see also Beckman, Jessen, & Ringen, 2013, for a related discussion.) The authors demonstrated that the lenis-fortis distinction in Swiss German is indeed reliably signaled by CF0. Most crucially, while both the fortis and the lenis stops show some commonalities in CF0 patterns largely comparable to the typical voiceless-related F0 perturbation effect, its detail revealed systematic CF0 differences due to the phonological opposition: CF0 starts relatively higher for the fortis than for the lenis, and the difference is maintained roughly halfway through the following vowel (rather than being localized to the vowel onset that an intrinsic CF0 would have shown). Furthermore, the Swiss German case presents another complexity in voicing contrast: the fortis stops /p t/ in quite a few words in the lexicon (e.g., names, loanwords and native words) are produced with aspiration, which is deemed to have gradually spread to more lexical items in Swiss German, possibly introducing a voiceless 'aspirated' category into the system. The authors reported that the voiceless stops that fall into this category show yet another distinct CF0 pattern with F0 being even higher than that of the fortis stop, though with some variation. The authors concluded that the phonetic manifestation of a potentially three-way stop contrast is primarily signaled by CF0, and it is in a strict phonetic sense independent of phonetic voicing. We take these findings as implying that the observed CF0 effect is not a mere reflection of low-level F0 perturbation that arises with the voicelessness of the stop,

but it may be internalized in, or controllable by, the phonetic grammar of the language as a phonetic parameter (CF0) that regulates voicing contrast (e.g., Kingston & Diehl, 1994). Swiss German may then be in its pathway on to a tonogenetic sound change.

Kirby (2018) also investigates the role of CF0 in voicing contrast in relation to VOT. Two tonal languages (Central Thai and Northern Vietnamese) and one non-tonal language (Khmer), all spoken in Southeast Asia, were studied. Each employs a notionally similar three-way stop contrast: (pre)voiced, voiceless unaspirated, voiceless aspirated. Exploring the effect of voicing contrast on CF0 in these languages provided a testbed for exploring two competing hypotheses. If CF0 is intrinsically driven by an automatic low-level phonetic effect, the degree of F0 perturbation reflected in CF0 would remain more or less comparable in both non-tonal and tonal languages. Alternatively, if the CF0 effect interacts with the phonological system of the language (i.e., under the speaker control), there will be differences between the tonal and the non-tonal languages, in such a way that, for example, its effect may be attenuated in tonal languages as it would otherwise blur the phonological tonal contrast. As for VOT distribution, results of this study in fact revealed a striking cross-linguistic similarity. The three-way voicing contrast in all three languages, tonal and non-tonal alike, is manifest clearly along the VOT continuum: prevoiced with mean negative VOTs from −60 to −74 ms; voiceless unaspirated with mean (positive) VOTs of 10–12 ms, and voiceless aspirated with mean VOTs of 75–94 ms. Crucially, however, the three languages did show noticeable differences in the extent to which CF0 contributes to the three-way contrast. The use of CF0 in conjunction with voicing contrast was found to be language-specific (and possibly speaker-specific) despite the cross-linguistic similarities in VOT distributions. For example, the non-tonal language (Khmer) showed the greatest magnitude and temporal extent of CF0, whereas the tonal languages (Thai and Vietnamese) revealed a generally attenuated CF0 effect, reflecting modulation of the voicing-related CF0s as a function of tonal context. Based on the cross-linguistic similarities in VOT and differences in CF0, the author concluded that CF0 is indeed something controllable by the speaker (thus being internalized in the phonetic grammar of the language) and the relation of the phonetic nature of the voicing contrast to the phonological system of the language may be illuminated by taking CF0 as an informative (but independent) complement to VOT, as was also suggested by Ladd and Schmid. An alternative explanation is that tonal and consonantal gestures overlap and therefore compete for control of F0 (Fowler & Brown, 1997; Pardo & Fowler, 1997). Listeners will then parse the effect (Fowler & Smith, 1986) so that non-tone language listeners will attribute the F0 different (appropriately) to the effect of the consonant voicing. No extra mechanism would then be necessary.

Al-Tamimi and Khattab (2018, this collection) also examine the effect of voicing contrast on CF0 in Lebanese Arabic along with VOT and other possible correlates. They show that while VOT may play a major role in distinguishing voiced and voiceless stops in both singletons and geminates, CF0 plays a further role in differentiating the voicing contrast in geminates, but not in singletons. To the extent that the effect holds, this suggests that the low-level effect may be suppressed or aug-

mented depending on its use in the language. In other words, the CF0 difference does not underlie the voicing contrast in the singleton but it supplements the voicing contrast in the geminate, which may be interpreted as suggesting that CF0 is under the speaker control to enhance the four-way contrast (Voicing × Quantity) in the language.

More broadly, the findings of these studies together imply that languages may employ a similar VOT pattern, but the combination of VOT and CF0 reveals cross-linguistic variation that might disentangle language specificity (and speaker specificity) from universally-applicable mechanisms that may involve phonetic voicing contrast in the world's languages.

4.2. Voice quality in relation to voicing contrast

Voice quality may play a role in making voicing distinctions, especially in languages that employ more than a two-way voicing contrast. Hindi, for example, employs a four-way voicing contrast—i.e., voiceless aspirated, voiceless unaspirated, voiced, and voiced aspirated, in which the fourth category 'voiced aspirated' is considered to have a laryngeal setting different from the other (typical) three laryngeal settings (Lisker & Abramson, 1964; Ladefoged & Maddieson, 1996). The 'voiced aspirated' category is generally known to be accompanied by breathy voice during the release phase (Dixit, 1989), which was taken by Lisker and Abramson as an indication that VOT was not appropriate for this category. In our call for papers for this special collection, we noted that our informal observation alluded to a possibility that the voiced aspirated in Hindi may be a combination of prevoicing during the closure and aspiration that follows without voice quality difference during the release, perhaps allowing positive and negative VOT in the same segment. While none of contributing papers has directly examined this possibility, Seyfarth and Garellek (2018, this collection) provide some relevant phonetic data in Yerevan Armenian (an Eastern variety) which is often described as employing voiced aspirated stops with 'breathy' voice quality to make a three-way contrast of, for example, /dʱ t tʰ/. Note, however, that Lisker and Abramson (1964) distinguished the three-way contrastive stops in Eastern Armenian in terms of VOT, which was mapped on to voiced, voiceless unaspirated and voiceless aspirated categories.

Seyfarth and Garellek provide evidence that the Armenian three-way stop contrast may indeed be generally distinguishable by VOT (negative, short-lag, long-lag VOT). Example tokens along with a waveform and its corresponding spectrogram (in Fig. 2 of their paper) further indicated that the voiced category does not seem to contain a substantial 'voiced' aspiration period, but it is accompanied by a relatively brief (partially devoiced) aspiration/release phase. If we consider the voiced duration as a negative VOT and the release phase as a positive VOT, the combination of the two may provide a unified metric for assessing voicing of voiced stops that may be accompanied by some degree of aspiration (whether voiced or voiceless). Hussain (2018) suggests that this combined metric may be more systematically applicable to assessing voicing distinction in Indo-Aryan and Iranian languages. In fact, example tokens provided in Fig. 3 of Hussain's paper indicate that the voiced aspirated stop in Indo-Aryan languages is characterized by both negative VOT and a following aspiration

portion which is much more prominent than the one observed in Yerevan Armenian. Thus, we suggest that the distinction between voiced categories especially between voiced unaspirated and voiced aspirated ones may be indeed better captured by the combination of the two measures, which may be integrated into one (extended) VOT metric (perhaps with VOT as a combination of the prevoicing interval and the aspirated interval after the release). Further analyses of existing data with such a metric will illuminate the extent to which a temporal extension of VOT would capture the multi-way voicing contrasts in these languages.

Setting aside the possibility of such an extended VOT, it is still of interest to look at the relationship between voice quality and the voicing distinction. Seyfarth and Garellek, in addition to VOT data, provide evidence that the voicing contrast is also clearly manifested in differences of voice quality (as reflected in $H1^*–H2^*$, an index of glottal constriction, and Cepstral Peak Prominence (CPP), an index of the noise). From the typological point of view, the results suggest that initial voiced stops in Yerevan Armenian are indeed breathy-voiced with clear phonation during the closure accompanied by a glottal spreading gesture which initiates during the closure and is maintained through the following vowel. The authors then compared the breathy-voiced stops in Yerevan Armenian to the voiced aspirated or breathy-voiced stops in Gujarati, another Indo-European language. It was noted that Yerevan Armenian voiced stops are provided with a relatively noisy (breathy) vowel without any evidence for the prevocalic voiced aspiration interval, whereas examination of available acoustic data in Gujarati (e.g., Esposito & Khan, 2012) indicates that the voiced counterparts in Gujarati are generally produced with an interval of voiced aspiration before a definable vowel onset, though with some variation among speakers. A substantial amount of breathiness during the following vowel was found to be present (as reflected in $H1^*–H2^*$) with the voiced aspirated stops in Gujarati, and the effect was more prevalent into the following vowel (compared to that in Yerevan Armenian). The authors concluded that the breathy voice quality does appear to play a role in voicing contrast in both languages, but the difference between the two languages may lie in the magnitude of glottal abduction gesture. Such a difference again can be taken to have been internalized in the phonetic grammar of the language, giving rise to cross-linguistic variation, which, as the authors suggested, characterizes the breathy-voiced stops in Yerevan Armenian as [b̤ d̤ g̤], but the voiced stops in Gurarati as [bʱ dʱ gʱ].

The results and discussion offered in the studies discussed so far lead us to some further thoughts on the use of VOT in relation to other phonetic parameters. VOT alone may suffice to describe the contrast within a language to the extent that the language employs an up to three-way laryngeal contrast in voicing as was originally proposed. But VOT was never intended to capture every aspect of the realization of voicing contrasts, so in order to understand cross-linguistic differences in phonetic properties of voicing contrast, it is necessary to look into other phonetic properties from multidimensional perspectives which VOT cannot capture. The multidimensional description may be phonologically redundant but phonetically informative in understanding universals and variation in voicing contrast in the world languages.
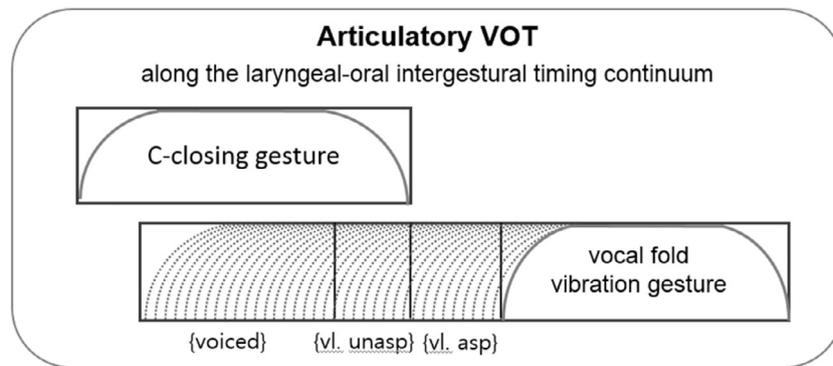
Fig. 3. The Articulatory VOT continuum in which the timing of vocal fold vibration gesture relative to C-closing gesture determines VOT distributions of both voiced and voiceless stops across languages.

5. Variation in phonetic implementation of phonological voicing contrast as a function of linguistic structure

Another topic for this special collection was the interplay between low-level phonetic realization of voicing contrast and higher-order linguistic structure. One possible theoretical consideration lies in the phonetics-prosody interface which informs how phonetic implementation of voicing contrast would be modulated by delimitative *vs.* cumulative functions of prosodic structure (e.g., prosodic boundary vs. prominence marking) of a given language (Keating, 2006; Shattuck-Hufnagel & Turk, 1996) and how it is related to the phonological system of the language (see Cho 2016 for a review).

Kim, et al. (2018, this collection) tackle this issue by examining the voicing contrast of initial stops in both trochaic and iambic words in American English. The prosodic-structural factors considered were boundary-induced domain-initial strengthening and prominence-induced strengthening (e.g., Cho & Keating, 2009; Cho, Kim, & Kim, 2017). The authors employed not only measures of VOT and voicing interval during closure, but also the Integrated Voicing Index (IVI), which may be seen as an extended version of VOT defined as a combined sum of a positive VOT and the voicing interval during the closure taken as a negative value. In this scheme, for a given token the IVI becomes negative when the voicing interval exceeds the short-lag VOT, but positive when the reserve is true. The authors suggested that this metric is particularly useful for assessing the degree of voicing of (phonologically) voiced stops that are often produced with both (short-lag) positive VOT and phonation interval during closure.

Their results suggest that boundary-related domain-initial strengthening conditions phonetic realization of initial stops (voiced or voiceless) in both trochaic and iambic words in a coherent way to increase 'voicelessness' of both phonologically voiced and voiceless stops. The unidirectional increase in voicelessness, despite their phonological opposition, showed no polarization of voicing contrast in the phonetic voicing (IVI) dimension, which the authors interpreted as suggesting an enhancement of a syntagmatic (CV) contrast at prosodic junctures. (Here by polarization the authors refer to a degree of separation (dispersion) in an opposite direction along the phonetic voicing dimension between the phonologically voiced and voiceless stops.) On the other hand, prominence-related strengthening (e.g., due to focus-induced nuclear pitch accent) indeed showed a pattern of polarization

of voicing contrast. Notably, however, it increased voicelessness for both voiced and voiceless stops, but the effect was far greater for the voiceless than for the voiced stops, dispersing the two stops farther away from each other along the voiceless IVI dimension (see Nelson & Wedel, 2017, for a related discussion). Moreover, the voiced stops showed its voicing value (IVI) centering near 'zero', often reducing the phonation interval during closure. Thus, the effect was not in the opposite direction (e.g., with voiced stops being more voiced) which would have been the case if the phonetic feature {voice} had been involved. The authors suggested that the prominence-related strengthening pattern is linked to a phonological voicing contrast with phonetic feature {vl. asp.} for voiceless stops and {vl. unasp.} for voiced stops, showing enhancement of language-specific phonetic features in reference to the higher-order prosodic structure. The results of this study therefore support the view that seemingly non-contrastive low-level phonetic voicing variation is indeed fine-tuned systematically by prosodic structure, and such phonetics-prosody interplay occurs in reference to language-specific phonetic representations that regulate the phonetic implementation of the phonological contrast in a given language.

The results in Kim et al. also have broader implications. Many of the studies presented in this special collection have investigated voicing contrast among stops in word forms produced in isolation or in sentence frames without looking into possible interactions between voicing contrast and prosodic strengthening effects. Further studies on the phonetics-prosody interface are called for. Particular attention should be paid to understanding how phonological voicing contrast is enhanced by its polarization pattern along the voicing (VOT) dimension and how it is further augmented by other phonetic parameters such as CF0 and voice quality within and across languages, which will again provide further insights into universals and variation of voicing contrast in the world's languages.

6. Sociolinguistics: language contact and levelling

Another possible area that we had hoped to cover in this special collection was 'sociophonetic' variation (Foulkes & Docherty, 2006) that may come from various social factors such as speaker gender, age, social class, dialect, language/dialect contact, speech style among many others. A particularly welcome contribution from the sociophonetic perspective

would be studies exploring how phonetic implementation of voicing contrast may be conditioned by these social factors, to what extent the observed variation may be understood as being rule-governed (or governed by the phonetic grammar of the language), and how this would inform linguistic modeling of the phonetics and phonology of voicing contrast in the language. We have not, however, received many contributions addressing these issues, except for one study which serves to reinforce our belief that further research on these matters is warranted.

Kleber (2018, this collection) investigates stop voicing contrast in two German varieties: Saxon German and Bavarian German. The two dialects have been known to show voicing contrast neutralization in initial position (in both Saxon and Bavarian) and in medial position (in Saxon), especially as far as VOT contrast is concerned. But given some recent evidence for dialect leveling due to the influence from standard German (which makes a clear phonetic voicing distinction between voiced and voiceless stops), Kleber particularly questions to what extent these two dialects show the dialect leveling in the form of a 'reversal' of the laryngeal neutralization. She carried out an apparent-time study by examining production of words produced in isolation and perception (2AFC) data obtained from two age groups (younger vs. older) in each dialect. In the production study, the phonetic voicing contrast of word-initial stops was assessed by VOT, which was normalized to the general speech rate (called pVOT). The phonetic voicing contrast of word-medial stops was assessed not only by pVOT, but also by another relative measure, called VCratio defined as a proportion of 'closure phase duration' to the entire V-C duration (where the closure phase duration referred to the interval from the onset of stop closure to the onset of the following vowel).

Results of the production study indeed provided some evidence for on-going dialect levelling. While both the older and the younger groups in each dialect manifest the stop voicing contrast in VOT whether initial or medial, the younger groups use VOT for voicing contrast to a greater extent than the older groups. The relative measure VCratio was also found to come into play in making the stop voicing contrast in medial position. The results of the perception experiment provided further evidence for dialect levelling. On the one hand, a perceptual trading relationship was observed between the two cues, VOT and VCratio, but in a direction that listeners do perceive the voicing contrast by means of a combination of VOT and VCratio. On the other hand, the younger listener groups in both dialects showed more reliance on VOT (with more categorical responses) than VCratio (with more gradual responses) in line with the expected use of VOT in standard German. The older listener groups relied more on VCratio, which reflected their own dialectal production pattern. With respect to the methodological issue, the author suggested that the 'quantity'-related VCratio should be considered in combination with VOT to capture phonetic manifestation of voicing contrast across different positions, and that their relative use in speech production and perception would illuminate possible on-going sound changes. But it remains to be seen to what extent the voicing contrast can be further illuminated through another means of voicing measure such as CF0 as discussed above, especially given that another dialect of German (Swiss German) appears to employ CF0

contrastively (Ladd & Schmid, 2018, this collection). In sum, this study demonstrates that some sound change that has been developed in a dialect-specific way may be reversed, which the author attributed to language contact and dialect levelling. It is hoped that this study sparks further cross-linguistic research on how voicing contrast may be adjusted by various other social factors as briefly mentioned above, and to what extent such an effect may be rule-driven in reference to linguistic systems of the languages. In particular, there is an opportunity arising from the availability of large naturalistic speech corpora for future research to investigate the extent to which VOT and other related properties of the voicing contrast are implicated in variability arising as a function of speech style – the dimension that has come to be seen as having a critical role in accounting for social-indexical variability in phonetic realization (e.g. Eckert 2000). This approach is likely to be enhanced by the refinement of new tools for the automatic measurement of VOT from large force-aligned corpora (e.g. Stuart-Smith, Sonderegger, Rathke, & Macdonald, 2015).

7. Articulatory mechanisms underlying voicing contrast

7.1. Ancillary articulatory manoeuvers in relation to voicing

In order to initiate and maintain voicing during closure there must be a critical difference in air pressure across the vocal folds (van den Berg, 1958; Westbury, 1983). If the intra-oral pressure is at a pressure similar to the subglottal pressure, there will be no airflow across the glottis, and hence no vocal fold vibration. The observed variation in VOT as a function of place of articulation for voiced stops in some of the languages including those Indo-Aryan languages and others (e.g., Russian) studied in this special collection may be taken to be attributable to differential intraoral pressure caused directly by the different sizes of the oral cavity behind the constriction (cf. Cho & Ladefoged, 1999). What is of particular interest then is how phonetic voicing (vocal fold vibration) may be facilitated by the speaker in order to improve the aerodynamic condition (i.e., achieving a sufficient transglottal pressure drop). One way of achieving this facilitatory goal is for the speaker to enlarge the oral cavity, which would effectively lower the intraoral air-pressure contributing to a transglottal pressure differential. An enlargement of the oral cavity may come to some extent passively due to compliance and plasticity of the vocal tract walls, but it may also involve a wide range of ancillary articulatory manoeuvers such as lowering the larynx, lowering the tongue and the jaw, raising the velum (to give more room), cheek bulging, and so on. It often involves slight opening of the velopharyngeal port, so that the resulting nasal venting lowers the intraoral pressure. Two studies in this special collection (Ahn, 2018a; Ünal-Logacev et al., 2018) have carried out articulatory experiments to explore the relationship between voicing and such supralaryngeal articulatory manoeuvers, and one study has tested the relationship between voicing and nasal venting (Kharlamov, 2018).

In an aerodynamic and EPG (electropalatographic) study, Ünal-Logacev et al. (2018, this collection) continue to explore voicing contrast of stops and affricates in Turkish, a 'true voicing' language, which revealed a substantial amount of prevoicing during closure of voiced obtruents. The authors were

particularly interested in exploring how the active maintenance of voicing during closure in a true voicing language like Turkish may be reflected in the linguo-palatal contact patterns of coronal obtruents /t tʃ/ *vs.* /d dʒ/. (For the aerodynamic measurement, a piezoresistive pressure transducer was used.) At the first analysis stage, the authors observed no relationship between the amount of linguo-palatal contact and aerodynamic measures in relation to voicing when the linguo-palatal contact was examined at one time point (i.e., a maximum contact point). But the authors found some meaningful relationships when the mutual dependence between the linguo-palatal contact and aerodynamic measures was considered through a Generalized Additive Mixed Model (GAMM). In particular, there was a linear relationship between the two measures (intraoral pressure and linguo-palatal contact) for the voiced stop /d/, but a non-linear relationship for the voiceless one. Crucially, as the linguo-palatal contact increased, the intraoral pressure increased for both voiced and voiceless stops, but the rate of increase in intraoral pressure was slower and more gradual for the voiced than for the voiceless stops, providing evidence for some kind of an articulatory strategy for facilitating phonation during closure. However, it may also be that the action of the vibrating folds tends to reduce the rate of pressure buildup directly. In any event, the authors suggest that the difference between the voiced stop /d/ and the other sounds may be understood as stemming from different laryngeal-oral coordination as a function of presence and absence of phonetic voicing, which is acknowledged to be subject to further corroboration. This study therefore takes one further step toward understanding the relationship between the laryngeal (voicing) contrast and the supralaryngeal articulation from the perspectives of the motor equivalence that may underlie voicing contrast across languages.

In an acoustic and ultrasound study, Ahn (2018a) investigates the relationship between voicing contrast and supralaryngeal articulation in English and Brazilian Portuguese which are generally categorized as an 'aspirating' language and a 'true voicing' language, respectively. It was hypothesized that if an enlargement of supralaryngeal cavity is primarily driven by improving aerodynamic conditions for phonation during closure, the supralaryngeal articulatory maneuver should be observable primarily in Brazilian Portuguese which employs phonation during closure, but such an effect is not expected to occur in English which often devoices the phonologically voiced stops in utterance-initial position. Alternatively, if the supralaryngeal articulatory difference is an integral part of the voicing distinction, the effects should be observable regardless of whether a language employs phonation during closure or not. Results showed that the voiced stops in Portuguese were indeed accompanied by a more advanced tongue root) as compared to their voiceless counterparts. This is consistent with the prediction that the advanced tongue root would effectively enlarge the supralaryngeal cavity (presumably by giving more room to the pharyngeal cavity) which would in turn help creating aerodynamic conditions for facilitating phonation during closure.

However, the phonologically voiced stops in English also showed similar articulatory behaviors as a function of phonological voicing contrast. This is an interesting finding because the phonological voice stops in English were generally devoiced, needing no articulatory adjustment for optimizing aerodynamic conditions for voicing. The only main difference observed between English and Portuguese was that the tongue root advancement in association with voiced stops were more consistently observed in Portuguese than in English. The difference, as Ahn suggests, reflects that Portuguese has a clear goal of phonation during closure (thus being more consistent) while English has no such clearly-defined goal of phonation (thus being more variable). Based on the similar articulatory manoeuvers in relation to voicing contrast in both 'true voicing' and 'aspirating' languages, the author drew an interim conclusion that supralaryngeal articulation is an integral part of laryngeal contrasts. To the extent that this effect holds across different languages, we can infer that the origin of such an ancillary articulatory manoeuver may have been phonetically grounded—i.e., in order to facilitate phonation during closure. It is then possible that the synergy of supralaryngeal and laryngeal gestures has been internalized in the phonetic grammar of the language, so that the system continues to employ it as an integral part of voicing contrast, despite there being no need for improving aerodynamic conditions any more due to the lack of voicing required for voiced stops in contemporary English. Such an explanation makes it more obvious that we should examine other contrasts for similar propensities that are not currently necessary for the language, but are either remnants of past organizations or seeds for future change.

Finally, in an acoustic and aerodynamic study, Kharlamov (2018, this collection) explores the phonetic nature of voiced stops in Russian, a 'true voicing' language. The phonetic data reported included prevoicing duration during closure, and aerodynamic measures of oral flow, oral pressure and nasal flow. What is particularly impressive about this study is that the phonetic data were obtained from a large pool of speakers (60 speakers) of Russian, a language whose detailed phonetic characteristics of voiced stops has been understudied. One of the important questions was to what extent prevoicing is accompanied by prenasalization, which may facilitate the aerodynamic condition (the critical transglottal pressure drop) required for phonation during closure. Results showed that prevoiced stops were indeed produced with a significantly larger amount of nasal venting as compared to their oral counterparts, but not as much as the nasal venting found for their phonemic nasal counterparts. The author interpreted the nasal venting of the voiced stops (or prenasalization) as resulting from an ancillary articulatory maneuver that involves regulating the opening of velopharyngeal port, which would in turn provide facilitative aerodynamic conditions for prevoicing.

In relation to this finding, one may assume that the nasal venting is a low-level phonetic effect observable across languages that may have been integrated as part of general automatic speech mechanisms that regulate phonation during closure. The data in Russian, however, allow for cross-linguistic comparisons (e.g., Solé, 2018; Solé & Sprouse, 2011) to test this possibility. The existing data in the literature indicate that the likelihood of nasal venting may vary systematically across languages not only between 'true voicing' languages (such as Spanish and French) and 'aspirating' languages (such as English), but also even among the true voicing languages. Importantly, Kharlamov shows that Russian employs the nasal venting to a greater degree than other 'true

voicing' languages discussed in the literature. This cross-linguistic comparison implies that the degree of nasal venting is employed under the speaker control and may be specified in the phonetic grammar of the language. Furthermore, the fact that the nasal venting in Russian varies in an increasing order from voiceless stops through voiced stops to nasals shows a systematic use of nasal venting within the language as well—i.e., the non-contrastive (low-level) nasality is fine-tuned according to phonological voicing contrast in Russian, and hence systematic prenasalization may not only facilitate voicing, but also manifest itself as another important phonetic (secondary) feature that may have been integrated into the phonetic system of the language (see Carignan, 2017, for another type of systematic covariation of phonetic features involving nasalization, breathiness and tongue height).

The studies discussed so far in this section provide important cases which specifically demonstrate how voicing contrast in true voicing *vs.* aspirating languages may be related to supralaryngeal articulatory adjustment, including tongue root advancement, modification of linguopalatal contact, and loosening of the velopharyngeal port, which are all assumed to facilitate aerodynamic conditions for voicing. These studies then leave more fundamental questions open as to universality of such an articulatory manoeuver (possibly driven, at least in part, by physiological biomechanic constraints imposed on the general human speech system) *vs.* its language-specificity which may be fine-tuned and thus internalized in the phonetic system (or grammar) of a given language. As we have discussed, these studies have provided some evidence for the systematic use of the potentially low-level effect in conjunction with linguistic contrast, which allows for testable hypotheses for the relation between voicing and supralaryngeal gestures. Elucidation of these issues requires further research on a wider range of languages with different laryngeal contrasts in languages. More specifically, it will be interesting to examine how the multiple laryngeal contrast in Indo-Aryan languages that distinguish voiced 'unaspirated' *vs.* voiced 'aspirated' stops will be manifest in the supralaryngeal articulation, how the difference in the assumed laryngeal settings between the voiced categories would be conjoined with any articulatory manoeuvers at the supralaryngeal level, and how the relationship between laryngeal contrast and supralaryngeal articulation may be internalized in the grammar of a given language. Similarly, as Ahn noted, more studies are needed to investigate the extent to which languages such as Mandarin that employ no actual phonation for laryngeal contrast would use supralaryngeal articulatory features as an integral part of the laryngeal contrast. More work is certainly called for to explore the relationship between the laryngeal and supralaryngeal articulation within and across languages.

7.2. Gestural approaches

VOT was originally proposed as an underlying variable that reflects phonetic consequences of voicing, aspiration and force of articulation which may be associated with differential laryngeal articulatory settings. In connection with laryngeal articulatory settings, a welcome theoretical consideration would have concerned how VOT may be modulated in the theoretical framework of Articulatory Phonology (Browman &

Goldstein, 1992). Goldstein (1992), for example, noted that "[t]he size (and timing) of a laryngeal gesture coordinated with an oral closure will determine the stop's voice-onset time (VOT). . ." (p. 212), implying that VOT is an output variable determined largely passively by the size of the glottal opening. This gestural approach further assumes that the timing of a laryngeal gesture with an oral gesture is already specified in the lexicon. In a similar vein, as briefly discussed above, Cho and Ladefoged (1999) (and Ladefoged & Cho, 2001) proposed 'Articulatory VOT' as a controllable variable whose model value is determined by the phonetic grammar of each language. (On a related point, Davidson (2018) observed that the proportion of partially phonated voiceless occlusions was similar for both aspirated and unaspirated realizations of underlying voiceless stops in American English (e.g., in stressed and unstressed syllables, respectively). These voiceless stops were also accompanied by a similar (so-called 'bleed') pattern of voicing that continued from the preceding vowel into the closure. Based on this observation, Davidson suggested that the start of the glottal opening gesture relative to the oral constriction gesture appears to be similar regardless of whether a voiceless stop is realized as aspirated or unaspirated. However, the difference in positive VOT between the voiceless aspirated and unaspirated stops is still needed to be determined in a language-specific way.) Setting aside the issue of whether the timing is directly specified in the lexicon or controlled by the language-specific phonetic grammar, both accounts, as they currently stand, appear to capture variation primarily for voiceless stops whose voicing is implemented along the positive VOT dimension.

For voiced stops, however, a basic assumption made in Articulatory Phonology (e.g., Browman & Goldstein, 1986, 1992) posits that vocal fold vibration occurs as a default mode in the absence of the laryngeal abduction gesture. In this regard, Kim et al. (2018, this collection) suggest that the voicing pattern of phonologically voiced stops in English may be in line with the 'default mode' assumption, if we assume that the voiced stops in English are unspecified for the laryngeal gesture, so that they are subject to passive voicing in a context in which voicing is facilitated (e.g. being flanked by vowels in prosodically weak position). Another way of understanding the phenomenon is to consider the phonetic feature {−spread glottis} which is articulatorily grounded as an underlying force. Given that the glottal adduction gesture as specified by {−spread glottis} provides a laryngeal setting for vocal fold vibration, its output may be contingent more on the contextual influence showing variation in VOT straddling the boundary between the negative and the positive dimension (as evident in Kim et al., 2018, and Ahn, 2018a, both in this collection). Either way, these possibilities may explain Davidson' (2016) finding that the amount of voicing during closure and its distributing pattern vary in phonetic implementation of voiced stops in English (e.g., 'bleed': voicing continues from the preceding vowel; 'hump': voicing occurs in the middle of the closure; 'trough': voicing discontinues in the middle of the closure and appears again before the release). But a further complication arises when we consider voicing realization patterns of voiced stops in 'true voicing' languages including those studied in this special collection. These languages show substantial voicing during closure for voiced stops, and the amount

of phonation duration is determined in a language-specific way, yielding variation across languages (as shown in Fig. 2). It may be that to account for the cross-linguistic variation, a mechanism needs to be devised in the framework of Articulatory Phonology that actively regulates the phonation gesture (see Davidson, 2018 for related discussion on possible gestural differences between aspirating and true voicing languages). However, it may also be that language-specific tuning parameters, such as strength of timing relationships, amplitude of segmental gestures, and interactions with prosodic control of F0, would provide a more automatic explanation.

The notion of 'Articulatory VOT' (Cho & Ladefoged, 1999; Ladefoged & Cho, 2001) allows for such an active control of phonation gesture. It assumes that the timing of vocal fold vibration is something that can be actively controlled in reference to the supralaryngeal gestural event. When this notion is extended to the voiced stops, we can provide a unified Articulatory VOT dimension, as schematized in Fig. 3, which is largely mapped on to the acoustic VOT dimension as suggested by Lisker and Abramson (1964) and Abramson and Whalen (2017). The voicing duration during closure may be relatively longer if the vocal fold vibration gesture (or voicing gesture) is timed early relative to the C-closing gesture (e.g., Lebanese Arabic as in Fig. 2), but it may be shorter if the voicing gesture is timed later relative to the C-closing gesture (e.g., Vietnamese as in Fig. 2). Thus, the variable timing latency of voicing gesture relative to C-closing gesture allows for cross-linguistic variation within the voiced category as marked by {voiced} in Fig. 3. If the voicing gesture is timed even later relative to C-closing gesture (after the release), the output enters the voiceless territory. The relative timing latency within the voiceless territory determines further whether the voiceless stop falls on the {vl. unasp} or {vl. asp} category, still allowing for within-category variation across languages.

The proposed Articulatory VOT account is not dissimilar from the basic notion of intergestural timing assumed in Articulatory Phonology, but it calls for refinement of the theory of Articulatory Phonology with regard to how it may account for variation in voicing that occur in the world's languages. The proposed Articulatory VOT scheme is rather sketchy, however, which leaves unsolved many other remaining issues that may concern any theories of voicing contrast from articulatory perspectives. Some of these issues are in need of further elaboration. First, given that the magnitude of the glottal opening gesture may influence the amount of VOT, it remains to be seen how it interacts with or influences the timing of the vocal fold vibration gesture relative to the C-closing gesture. Second, a more sophisticated mechanism is needed to account for the difference between voiced unaspirated and voiced aspirated stops that occur in Hindi and other Indo-Aryan languages. Two separate gestures may be needed for voiced *aspirated* stops which show a combination of voicing and aspiration. It may be possible to specify two gestures: vocal fold vibration (voicing) gesture and glottal spreading gesture. Given that voicing continues with aspiration in the voiced aspirated stop, the part of the glottal spreading gesture may be superimposed on the voicing gesture (which induces some degree of breathy voicing caused by vocal fold vibration during partial laryngeal adduction), but they both should be timed relative to C-closing gesture, resulting in cross-linguistic variation. On the one hand, if the glottal spreading gesture is timed earlier relative to the release of C-closing gesture, voicing murmur may arise as in [b̤ d̤ ɡ̤] in Yerevan Armenian. On the other hand, if the glottal spreading gesture is aligned with the release of C-closing gesture, it may result in voicing patterns appropriate for voiced aspirated stops [bʱ dʱ ɡʱ] found in Indo-Aryan languages. Finally, if the glottal spreading gesture straddles the release of C-closing gesture, it may result in voiced stops which are characterized by both voicing murmur during closure and the following voiced aspiration period, a typologically possible category that has not been attested. (See Ahn's PhD dissertation (Ahn, 2018b) for related discussion on how the difference between the voiced aspirated and voiceless aspirated stops may be explained by a different magnitude of the glottal opening gesture).

Another remaining issue concerns how within-language variation as a function of prosodic structure may be understood in terms of intergestural timing as discussed in Kim et al. (2018, this collection). For example, different types of prosodic strengthening (boundary-induced vs. prominence-induced) may result in re-organizing intergestural timing relationship, which also must be fine-tuned further in reference to types of linguistic contrast that underlie different prosodic strengthening patterns. Prosodic strengthening may also influence the spatial dimension both at the laryngeal and the supralaryngeal level, so that it may often augment the spatial displacement that may be associated with the articulatory target of the underlying gesture (e.g., Byrd & Saltzman, 2003; Cooper, 1991; Cho, 2006; Cho & Keating, 2009; de Jong, 1995). For example, Cooper (1991) observed an expansion of glottal opening for voiceless stops in utterance-initial position relative to utterance-medial position (cf. Pierrehumbert & Talkin, 1992), showing a kind of domain-initial strengthening of laryngeal gesture. These prosodic-structurally related possibilities together warrant further studies on languages which employ 'true voicing' and/or 'voiced aspiration' to examine the extent to which voicing-related gestures (vocal fold vibration gesture and glottal spreading gesture) may be strengthened in the spatial domain as well as in the temporal domain relative to the supralaryngeal articulation, and how the resulting pattern would relate to enhancement of linguistic contrasts such as paradigmatic vs. syntagmatic ones. In recent years, we have seen some advancement in understanding mechanisms of regulation in speech in relation to linguistic structure and physical control system (e.g., Mücke, Hermes & Cho, 2017) in general, and how the spatio-temporal realization of articulatory gestures may be modulated by prosodic factors including those that may induce different kinds of prosodic strengthening (see Krivokapić, in press, for a review). It remains to be seen how the detail of prosodically-structurally conditioned articulatory variation in relation to voicing contrast can be accounted for by the gestural approaches, which will help us understand universal vs. language-specific articulatory underpinnings of voicing contrast in the world's languages.

8. Conclusion

We have reviewed 11 studies (including Abramson & Whalen, 2017) on voicing contrast in 19 languages (21 varieties with three German dialects) that have contributed to the

special collection marking 50 years of research on Voice Onset Time. We have also provided further discussion of the insights that these studies have brought together to the field, which illuminates the typology of laryngeal contrast in voicing with some new perspectives on universals and variation in laryngeal contrast in the world's languages. These studies showed ample evidence for the usefulness of VOT as a simple metric for assessing laryngeal contrast in voicing across many languages, especially for those languages that employ a binary or a three-way voicing contrast, as was originally envisaged by Lisker and Abramson (1964), and reiterated by Abramson and Whalen (2017, this collection). The studies also demonstrated that languages may choose their modal VOT value for each phonetic category, which should be separately specified in the phonetic grammar of each language, yielding cross-linguistic variation as was suggested by Cho & Ladefoged (1999). But the cross-linguistic data indicated that the scope of variation should still be determined within a permissible window allowed by each of the universally available phonetic categories, such as voiced, voiceless unaspirated, and voiceless aspirated largely in line with proposals made by Keating (1984, 1990a). It is probably this sort of informativeness of VOT that has long driven the field to resort to VOT as a first estimate of voicing contrast within and across languages.

Despite the descriptive power of VOT, it was never expected to account for all phonetic aspects of the voicing distinction. We have seen further evidence that the phonetic and phonological nature of laryngeal contrast in voicing may never be complete without looking into other phonetic properties that may co-occur with or replace VOT. In particular, the multi-way laryngeal contrast in many Indo-Aryan languages poses a challenge for the use of VOT, and necessitates either an extension of VOT or an exploration of other possible correlates such as CF0 (consonant-induced F0 effect on the vowel) or voice quality. The need also extends to other cases in which the difference in VOT is no longer distinctive as in Swiss German. Moreover, even when VOT plays a clear role in making voicing distinction, co-occurrence of other complementary cues such as CF0 and voice quality may further inform how voicing contrast in these languages may have evolved multi-dimensionally and possibly on different pathways in relation to other higher-order linguistic structure (such as tonal contrast or prosodic structure) and extralinguistic structure (as evident in linguistic levelling through dialect contact), engendering linguistic arbitrariness within universals. Some studies further discussed the usefulness of the multi-dimensional approach by examining the relation of laryngeal contrast to supralaryngeal articulation, which is phonetically grounded (for facilitating phonation during closure), but may eventually be internalized as an integral part of voicing contrast in the sound system of the language. Finally, we discussed the notion of Articulatory VOT and gestural approaches which could best account for variation in VOT within and across languages by modulating intergestural timing between vocal fold vibration gesture and oral constriction gesture.

To conclude, the contributing studies in this special collection have provided new insights into the phonetic and phonological nature of laryngeal contrast in the world's languages. They have, of course, left us with new questions to be answered, but with useful pointers to where the field may be best directed. It is our hope that this impressively extensive research collection on voicing contrast in 19 languages will lead to proliferation of further work on laryngeal contrast in voicing in the world's languages uncovering what underlies universals and variation in VOT and beyond it.

References

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica, 34*, 295–303.

Abramson, A. S. (1995). Laryngeal timing in Karen obstruents. In F. Bell-Berti & L. Raphael (Eds.), *Producing speech: Contemporary issues. For Katherine Safford Harris* (pp. 155–165). Woodbury, NY: American Institute of Physics Press.

Abramson, A. S., & Whalen, D. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63*, 75–86.

Ahn, S. (2018a). The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics, 71*, 451–467.

Ahn, S. (2018b). *The role of tongue position in voicing contrasts in cross-linguistic contexts* PhD dissertation. New York University.

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics, 71*, 306–325.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics, 66*, 120–144.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics, 49*, 259–284.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*, 149–180.

Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics, 63*, 87–105.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics, 37*(4), 466–485.

Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Papers in laboratory phonology VIII* (pp. 519–548). Berlin & New York: Mouton de Gruyter.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass, 10*(3), 120–141.

Cho, T., & Ladefoged, L. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207–229.

Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics, 64*, 71–89.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47.

Coetzee, A. W., Beddor, P., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics, 66*, 185–216.

Cooper, A. M. (1991). *Glottal gestures and aspiration in English* PhD dissertation. Yale University.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50.

Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association, 48*(3), 331–356.

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 97*, 491–504.

Dixit, R. P. (1989). Glottal gestures in Hindi plosives. *Journal of Phonetics, 17*, 213–237.

Docherty, G. (1992). *The timing of voicing in British English obstruents.* Berlin; New York: Foris.

Eckert, P. (2000). *Linguistic variation as social practice.* Oxford: Blackwell.

Esposito, C. M., & Khan, S. (2012). Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association, 42*, 123–143.

Foulkes, P., & Docherty, G. J. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*, 409–438.

Fowler, C. A., & Brown, J. M. (1997). Intrinsic f0 differences in spoken and sung vowels and their perception by listeners. *Perception and Psychophysics, 59*, 729–738.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldstein, L. (1992). Comments on chapters 3 and 4. In G. Docherty & D. R. Ladd (Eds.), *Laboratory phonology, Vol. 2: Gesture, segment, prosody* (pp. 120–124). Cambridge: Cambridge University Press.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America, 125*, 425–441.

Hussain, Q. (2018). A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics, 71*, 284–305.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76–90.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286–319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in Honor of Peter Ladefoged* (pp. 115–132). Orlando FL: Academic Press.

Keating, P. A. (1990a). Phonetic representations in a generative grammar. *Journal of Phonetics, 18*, 321–334.

Keating, P. A. (1990b). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 451–470). Cambridge: Cambridge University Press.

Keating, P. A. (2006). Phonetic encoding of prosodic structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167–186). New York and Hove: In Macquarie Monographs in Cognitive Science Psychology Press.

Kharlamov, V. (2018). Prevoicing and prenasalization in Russian initial plosives. *Journal of Phonetics, 71*, 215–228.

Kim, S., Kim, J., & Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics, 71*, 65–80.

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

Kingston, J. (2007). Segmental influences on f0: Automatic or controlled? In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes, volume 2: Experimental studies in word and sentence prosody* (pp. 171–201). Berlin: Mouton de Gruyter.

Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.). *Blackwell companion to phonology* (V. 4, pp. 2304–2334). Oxford, UK: Blackwell Publishing.

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

Kleber, F. (2018). VOT or quantity: what matters more for the voicing contrast in German regional varieties? Results from apparent-time analyses. *Journal of Phonetics, 71*, 468–486.

Kohler, K. J. (1982). F0 in the production of fortis and lenis plosives. *Phonetica, 39*, 199–218.

Krivokapić. J. (in press). Prosody in articulatory phonology. In Shattuck-Hufnagel & J. Barnes (Eds.), Prosodic theory and practice. MIT Press.

Ladd, D. R., & Schmid, S. (2018). Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics, 71*, 229–248.

Ladefoged, P., & Cho, T. (2001). Linking linguistic contrasts to reality: The case of VOT. In N. Gronnum & J. Rischel (Eds.), *Travaux Du Cercle Linguistique De Copenhague, vol. XXXI. (To Honour Eli Fischer-Forgensen.)* (pp. 212–223). Copenhagen: C.A. Reitzel.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell Publishers.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.

Löfqvist, A., Baer, T., McGarr, N., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*, 1314–1321.

Nelson, N. R., & Wedel, A. (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics, 64*, 51–70.

Mücke, D., Hermes, A., & Cho, T. (2017). Mechanisms of regulation in speech: Linguistic structure and physical control system. *Journal of Phonetics, 64*, 1–7.

Pardo, J. S., & Fowler, C. A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception and Psychophysics, 59*, 1141–1152.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–117). Cambridge, UK: Cambridge University Press.

Seyfarth, S., & Garellek, M. (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics, 71*, 425–450.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.

Solé, M.-J. (2018). Articulatory adjustments in initial voiced stops in Spanish, French and English. *Journal of Phonetics, 66*, 217–241.

Solé, M.-J., & Sprouse, R. (2011). Voice–initiating gestures in Spanish: Prenasalization. In M. Wei & E. Zee (Eds.), Proceedings of the 17th international congress of phonetic sciences (pp. 72–75). Hong Kong, China.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., & Macdonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology, 6*(3–4), 505–549.

Ünal-Logacev, Ö., Fuchs, S., & Lancia, L. (2018). A multimodal approach to the voicing contrast in Turkish: Evidence from simultaneous measures of acoustics, intraoral pressure and tongue palatal contacts. *Journal of Phonetics, 71*, 395–409.

van den Berg, J. (1958). Myoelastic theory of voice production. *Journal of Speech and Hearing Research, 1*, 227–244.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America, 73*(4), 1322–1336.