HHS Public Access

Author manuscript

J Phon. Author manuscript; available in PMC 2018 July 01.

Published in final edited form as:

J Phon. 2017 July ; 63: 75–86. doi:10.1016/j.wocn.2017.05.002.

Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions

Arthur S. Abramson[a] and **D. H. Whalen**[a,b,c,*]

[a]Haskins Laboratories, 300 George St., Suite 900,NewHaven, CT 06511,United States

[b]Program in Speech-Language-Hearing Sciences, Graduate Center, City University of New York, 365 5th Ave., New York, NY 10016, United States

[c]Department of Linguistics, Yale University, PO Box 208366, New Haven, CT 06520-8366, United States

Abstract

Just over fifty years ago, Lisker and Abramson proposed a straightforward measure of acoustic differences among stop consonants of different voicing categories, voice onset time (VOT). Since that time, hundreds of studies have used this method. Here, we review the original definition of VOT, propose some extensions to the definition, and discuss some problematic cases. We propose a set of terms for the most important aspects of VOT and a set of Praat labels that could provide some consistency for future cross-study analyses. Although additions of other aspects of realization of voicing distinctions (F0, amplitude, duration of voicelessness) could be considered, they are rejected as adding too much complexity for what has turned out to be one of the most frequently used metrics in phonetics and phonology.

Keywords

VOT (Voice Onset Time); voicing; consonants; stops; duration

1. Introduction

Just over fifty years ago Leigh Lisker and Arthur S. Abramson proposed, with acoustic data from 11 languages, that the timing of glottal pulsing relative to supraglottal articulation would account for the great majority of homorganic consonantal distinctions traditionally said to depend on voicing, aspiration, “tensity,” and the like (Lisker & Abramson, 1964). Supporting data were furnished in early studies of laryngeal behavior in stop consonants (Lisker, Abramson, Cooper, & Schvey, 1969; Sawashima, Abramson, Cooper, & Lisker,

*Correspondence to: Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, United States. Tel: +1-203-865-6163; fax +1-203-865-8963. whalen@haskins.yale.edu (D. H. Whalen).

1970), speech acoustics (Lisker & Abramson, 1964), and perception (Abramson & Lisker, 1965; Lisker & Abramson, 1970).

Voicing distinctions among stops are, of course, quite common in the world's languages, and they have been encoded in alphabetic and syllabic scripts for millennia. The acoustic realizations were not observable, however, until the modern era (cf. Tillmann, 1995). Some of these led to precursors of the VOT concept. Rousselot (1897–1908, p. 456) noted a difference in the timing of voice onset between French and German plosives, but he did not elaborate on that distinction further. Panconcelli-Calzia (1924, p. 45) observed that the difference in timing between release and voicing varies, and that a value that would count as "voiced" in one language would be "voiceless" in another. There seems to have been essentially no further early work in this vein. Somewhat later, at Haskins Laboratories, manipulation of noise and "F1 cutback" (starting F1 later than F2 and F3) during transitions (Liberman, Delattre, & Cooper, 1958) was found to signal a shift from voiced to voiceless in consonant voicing. While this was a partial precursor to the concept of VOT, those authors used only a single duration of noise and F1 cutback, so the important temporal dimensions remained to be explored.

1.1 First definition

Since the 11 languages that Lisker and Abramson (1964) wished to examine all had voicing distinctions in word-initial position but not necessarily elsewhere, the research in the foundational article was limited to word beginnings (Lisker & Abramson, 1964). Thus it was that the name voice onset time (VOT) was given to the dimension under consideration. It must be emphasized, however, that even then the study included word-initial stops in sentence-medial positions. In later sections, we will add discussion on the role of laryngeal timing in medial and final positions.

In the original paper, VOT was defined as the temporal relation between the moment of the release of the stop and the onset of glottal pulsing (see Fig. 1). That is, the orientation was toward speech articulation. If the pulsing started in the consonantal closure and continued to the stop release and beyond, it was treated as an instance of voicing lead. If it started after the release, it was called voicing lag. With the stop release treated as a reference point at 0 ms, measurements of voicing lead had negative values, while those of lag had positive values. The measurements were normally made on wideband spectrograms (with better time resolution than narrow-band ones), although they were sometimes cross-checked in waveforms. Thus, for example, a "fully voiced" stop would show voicing lead through much or all of its closure and through its release into whatever follows (Fig. 1, top panel); a "voiceless unaspirated" stop would have a VOT of 0 ms or a very short lag with very brief aspiration in the lag (Fig. 1, middle panel); a "voiceless aspirated" stop would have a somewhat larger amount of voicing lag with aspiration throughout (Fig. 1, bottom panel). Cho and Ladefoged (1999) argue that the boundary must be defined on a language-by-language basis. Aspiration is presumably a result of a somewhat open glottis. Note that this means that the presence of a small amount of aspiration is not sufficient to (perceptually) mark a stop as "aspirated." In the matter of long voicing leads, the trans-glottal air pressure must be great enough to allow glottal pulsing for the duration of the stop closure. One way

of delaying pressure equalization and the consequent stoppage of voicing is through enlargement of the supralaryngeal cavities (Westbury, 1983). In contrast with citation forms and very deliberate running speech, after a voiceless segment glottal pulsing may have a delayed onset in the following closure of a "voiced" stop or, especially in English (Baran, Laufer, & Daniloff, 1977; Davidson, 2016) no pulsing at all or short stretches of pulsing followed by silence until the moment of consonant release. (A detailed list of such combinations can be found in Docherty, 1992, p. 197.)

As noted below (Section 2.), it is difficult to distinguish between the noise of a release burst and the noise of aspiration, so that some prevoiced stops with an intense burst may seem to have both prevoicing and aspiration. In Fig. 2, the second author's production of a prevoiced [ga], shows such a case. This was a deliberate use of the variability that is found for English (prevoicing rather than simultaneous release), both in the original article and many other studies. The burst is marked with gray shading. The prevoicing declines in amplitude almost to silence before the release, and the burst is quite voiceless. However, in this example the burst is followed by a resumption of glottal pulses that are at first rather weak with perhaps some turbulence intermingled. This should count by our definition as −134ms VOT (the duration of the closure); the voicelessness after release could be measured for other purposes (such as estimating boundary strength), but this token is prevoiced.

In intervocalic position, the measurement of prevoicing (or continuous voicing) is somewhat more complicated. A small amount of voicing at the beginning of intervocalic closures was referred to as "edge vibrations" in the original article (Lisker & Abramson, 1964, p. 417). Such examples can be seen in Davidson (2016, p. 43), labeled as "bleed"; nonetheless, she only classifies one of the four patterns with voicing in the closure as "negative VOT." Her criterion is that "partial voicing" occurs when 10–90% of the closure is voiced. This makes it unclear how to categorize the VOT in those cases, on the assumption that we should be able to provide a VOT in all cases. Her "hump" pattern (her Fig. 5d) represents a case in which the closure begins without voicing, has voicing for a portion of the closure but returns to silence before release. It would be classified as partially voiced, but it is unclear what the VOT would be. The natural changes in voicing amplitude during a closure, as intraoral air pressure builds, combined with the necessarily noisy release burst lead to this disregard of voicing in the closure and an increase in the number of examples that are classified as having a positive VOT.

1.2 Some limits on application

There are languages in which VOT alone will not handle every one of the distinctions between classes of stop consonants that might be seen as part of a voicing contrast. An additional phonetic dimension is relevant in such instances. An example is Hindi with its voiced aspirated stops (Lisker & Abramson, 1964, pp. 397–398). Thus, it has four categories, as illustrated here with just the labial place of articulation.

LABIAL b b^h p p^h

That is, along with some other languages of the Indo-Aryan family, it has not only the stops that are well differentiated by VOT but also voiced aspirated stops, which are not distinguished from voiced stops by VOT. Rather, the two voiced categories are distinguished

by voice quality (phonation type) (Dixit, 1975, 1979; Dutta, 2007; Shimizu, 1996). In the voiced aspirates, the lag before the onset of modal phonation is filled with murmur (breathy voice). In these languages, however, VOT does distinguish the voiceless unaspirated and voiceless aspirated stops as well as the latter two from the two voiced categories, which themselves are differentiated by another property, phonation type. This implies that voiced aspirates should be assigned a single VOT value, one that is negative, with the caveat that the VOT does not encompass the entire voicing distinction.

Alternative interpretations for the Hindi voiced aspirates have been proposed. Davis (1995) claims that "noise offset time" is sufficient to divide the stops, and that murmur is not always present and thus not distinctive. In their analysis of the similar stops in Gujarati, Mikuteit and Reetz (2007) claim more directly that voicing is realized as pulsing during closure and that the voiceless signal after release ("After Closure Time" or "ACT") plus superimposed aspiration ("SA") on the following vocalic segment provides a four-way distinction consistent with two (potentially underspecified) phonological features. The distinction between SA and murmur is defined as "some part of what is commonly known as breathy voice" (p. 250). It is unclear how this is to be distinguishd from murmur. The overall means in ACT duration are indeed longer for the voiced aspirates than for the inaspirates, but there is no indication of the amount of overlap in the various tokens. The authors collapse across aspiration in their reporting of SA durations (p. 274), so it is not possible to tell whether that component is the major contributor to the distinction. Although the situation is somewhat complicated, there is still validity in the original designation of the voiced aspirates being outside the domain of VOT.

Another example is Korean. This is a three-category language, but in utterance-initial position all three are voiceless. From the point of view of VOT one could label them: I. unaspirated, II. slightly aspirated, and III. heavily aspirated, yielding successively larger average VOT values for isolated words, although the ranges of distribution of the data show rather poor discrimination for the first two categories (Lisker & Abramson, 1964, Table 9 and Figs. 5, 6, and 7). VOT does differentiate all three categories quite well, however, when Category II is in non-initial position preceded by voicing; that is, the closure is excited by audible glottal pulses. Note, however, that it is the "slightly aspirated" category that is fully voiced in medial position. Perceptual data obtained with synthetic speech (Abramson & Lisker, 1972) give general support to these acoustic findings. All of this has to be considered against a background of controversy. Earlier descriptions of Korean stops were based, of course, on the practical phonetics of fieldwork (e.g., Martin, 1951). Such terms as "tense" and "lax" were used in those days without any convincing phonetic evidence as to their meaning. In his study of acoustic amplitude, air pressure, and other properties, Chin-Wu Kim (1965) claims support for tensity as a primary differentiator of Korean stops. Curiously enough, in a later study (1970), he ignores this claim of his and presents data on varying glottal width for a theory of aspiration with special reference to the stops of Korean. This, by the way, is in agreement with the description in terms of degree of aspiration in the Korean sample published by the International Phonetic Association (Lee, 1993); this description, however, makes no reference to the instrumental studies of the topic. In any event, Korean is a language in which VOT does not tell the whole story, except in certain non-initial contexts. Some other property is also involved, perhaps in the behavior of laryngeal musculature (e.g.,

Kagaya, 1974). Indeed, it has been claimed that certain dialects of Korean show an emerging tonal system that helps with the case of VOT overlap in initial position (M.-R. Kim, Beddor, & Horrocks, 2002; Silva, 2006). Kang (2014) found that this was the case for younger female speakers of Seoul dialect. Cho, Jun and Ladefoged (2002) report that the three-way Korean stop contrast may be further reflected in other phonetic parameters, such as F0, H1-H2, acoustic burst energy, intraoral pressure and airflow, which may support the three-way stop contrast perceptually. Nonetheless, it is certainly possible to measure VOT in all Korean stops, again with the caveat that it is not quantifying all aspects of voicing.

Thus, contrary to Kim (1965, p.359) it was never claimed, either in the founding article (Lisker & Abramson, 1964) or thereafter, that VOT is universally relevant for all distinctions of stop categories. Indeed, exceptions noted here were mentioned from the start.

The pre-aspirated stops of such languages as Icelandic, Swedish and Tarascan (Silverman, 2003) are also not well represented by VOT. For example, in the pattern reported for Swedish, there is a brief (ca. 16 ms) period of aspiration followed by a long, silent occlusion (ca. 69 ms) (Helgason, 2002, p. 121). In Scots Gaelic, the pre-aspiration appears to be more equivalent in duration to the closure (ca. 80 ms for both, extrapolating from the figures) (Ladefoged, Ladefoged, Turk, Hind, & Skilton, 1998, p. 10), but, for their speakers at least, there is approximately 35 ms of aspiration after the release as well (p. 11). Indeed, Nance and Stuart-Smith (2013) measure both duration of preaspiration and VOT. In general, the total duration of the devoicing associated with the aspirated stop extends from the previous vocalic segment (the pre-aspiration) to the onset of voicing in the following vocalic segment. This duration is of interest in itself, but it requires the use of a different metric (duration of devoicing) that is also present in other languages (English included) that have silent closures. Further, to our ears, many examples of the “aspiration” in preaspirates are fricatives, not aspiration, further complicating the issue. Indeed, preaspiration can be seen phonetically as voicing offset, i.e., devoicing with aspiration in the latter part of the vowel preceding the stop closure. Assigning this kind of aspiration to the syllable-final stop closure is a phonological decision based on its distributional limitation to a particular set of following stops. Such a language then could even have a four-way contrast between voiceless unaspirated stops, voiceless aspirated stops, voiceless preaspirated stops, and voiced stops. Thus the treatment of preaspiration requires more than can be incorporated into a simple VOT value.

Two other manners of articulation of stops can be described with VOT, ejectives and implosives. Ejectives are necessarily accompanied by a positive VOT, but VOT does not necessarily distinguish ejectives from aspirated stops (see data in Ladefoged & Cho, 2001). Ejectives have two closures, one glottal and one oral, and the glottal is released after the oral, so no aspiration is present, a further difference from typical long VOT stops. Implosives are necessarily prevoiced or accompanied by breathy voice, and, again, are not distinguished from voiced stops by VOT (e.g., Bennett, 2010).

Even in stop systems well differentiated by VOT, other properties have been found to play an ancillary role. The strength of prosodic boundary has been shown to induce gradience in the VOT of Korean stops (Cho & Keating, 2001). Increases in the amplitude of the

aspiration can lead to more voiceless judgments in perception (Repp, 1979), as if the VOT were longer. More surprisingly, the fundamental frequency of the voice upon release of the consonant closure tends to be higher in the presence of voicelessness (Hombert, Ohala, & Ewan, 1979). This has been attributed to contraction of the cricothyroid muscle of the larynx to help cut off pulsing of the vocal folds (Löfqvist, Baer, McGarr, & Story, 1989). Indeed, in perceptual experiments with synthetic speech (Abramson & Lisker, 1985; Haggard, Ambler, & Callow, 1970), it has been shown that such perturbations of F0 can affect the VOT value at which the categories are distinguished. Later, more elaborate perceptual experiments (Whalen, Abramson, Lisker, & Mody, 1990, 1993) reaffirmed that finding and showed effects of F0 even for otherwise unambiguous VOT values. As far as we know, the power of F0 as a cue to voicing in natural running speech has not been established, though in Spanish and English, F0 does begin at a higher value for voiceless stops regardless of VOT (Dmitrieva, Llanos, Shultz, & Francis, 2015). However, including such perceptually relevant measures would require that VOT included amplitude and F0 values as well, making it more than just a temporal measure and eliminating its useful simplicity.

1.3 Phonological value

The concept of laryngeal timing as a physiological mechanism (Abramson, 1977; Abramson & Lisker, 1969) might be expected to have a simplifying and clarifying effect (Lisker & Abramson, 1971). Of course, the notion of a temporal dimension as a “distinctive feature” may be rather startling for many phonologists who are reluctant to hug the phonetic ground even while using phonetic terms in their analyses (Lisker & Abramson, 1987). Keating (1984) uses the patterns of VOT in different segmental environments to uncover whether voicing or aspiration is distinctive for a particular language. VOT has figured in Optimality Theory as a component of a constraint that defines contrast (e.g., Flemming, 2013, pp. 47–48; Steriade, 2000, pp. 332–333). Some phonological features have been proposed that incorporate VOT (Gallagher, 2011, 2015). Most phonologists, however, take VOT as an aspect of the realization of more standard features such as +/− voice (e.g., Wetzels & Mascaró, 2001) and/or +/− spread glottis (e.g., Beckman, Jessen, & Ringen, 2013). There is also evidence that the phonological status of the VOT categories affects implementation at prosodic boundaries (Cho, Lee, & Kim, 2014; Cho & McQueen, 2005).

Mikuteit and Reetz (2007) propose a radically underspecified approach (again without a time component) in which voicing during closure indicates +voice, aspiration indicates +spread glottis, and negative values of those features are not specified. Thus voiceless aspirates are only indicated as +spread glottis, while voice inaspirates have no specification for either feature. A language like English has a spread glottis distinction, not a voicing distinction; one has to assume that the voicing of non-aspirated stops in intervocalic position would be an allophonic rule or a default setting. While this model provides a concise description of their own data, it is less clear how it applies in cases where the distinction between simple release and aspiration is not clear (e.g., Korean) and the pre-aspirates. Further work in phonology seems to be necessary.

An analysis with inherent time value is, however, explicit in Articulatory Phonology (e.g., Browman & Goldstein, 1992). In that theory, most languages with two voicing categories

use the presence or absence of a glottal gesture for the distinction, with differences in timing and/or magnitude accounting for different VOT patterns (Goldstein & Browman, 1986). A similarly articulatory-based analysis is found in Ladefoged and Cho (2001). Such an analysis would lead us to want to measure the duration of the glottal gesture itself instead of, or in addition to, the VOT. This would presumably show greater consistency in the gesture across place of articulation than VOT does, given that longer VOTs with more posterior places (e.g., Lisker & Abramson, 1967) are accompanied by shorter closure durations. However, there are many factors involved in assessing a glottal gesture, and these simple acoustic measurements are inadequate to the task. More direct and complete assessments of the glottal gesture are needed.

Voicing categories in many of the world's languages are captured as differences in VOT, while some categories (voiced aspirates, Korean lax stops) are not. Distinctive features can handle such division, as they can with any set of categories. More detailed analyses, those that attempt to take the timing information of VOT into account, show promise but are not yet fully developed.

2. Modified definition

From the start, investigators have given attention to laryngeal timing not only in utterance-initial position but also in utterance-medial position including word-medial position. Now it must be remembered that voice is not just present for phonemic distinctions; in fact it is the normal carrier of speech signals. It is interrupted wherever the language being spoken requires silence or the presence of some other excitation, normally some kind of noise. Thus, in an isolated utterance of the word *books* the labial closure of the initial /b/ is likely to be silent with a voiceless release burst of the closure followed almost immediately by the onset of voicing. The voicing ceases for the final /ks/ cluster, though there is evidence of some voicing after closure. If, however, the utterance is *two books*, the closure and release of the word-initial /b/ are typically fully voiced by the glottal pulses proceeding unbroken from the preceding vowel of *two* (see Fig. 3). Thus, the voicing lead can be associated with the closure, just as it can in absolute initial stops, and thus a negative value in milliseconds can be measured for it. In the foundational article (Lisker & Abramson, 1964) and thereafter this was simply called "unbroken voicing." (This was labeled "connection voicing" or CVO by Mikuteit & Reetz, 2007.) Likewise, to continue with English, the word-pair *tugging* and *tucking*, the fully voiced /g/ is characterized by unbroken voicing and the unaspirated /k/ by a measurable gap in voicing, albeit with partial voicing and a noise-filled "closure" (see Fig. 4). Such incomplete closures are common in running speech (Crystal & House, 1982; Shockey, 2003, p. 27). Thus, the concept of voice timing still prevails. The acronym VOT might still be appropriate (even though the "onset" is dubious) to keep the terminology consistent with other measures. Alternatively, MVOT might indicate "medial voice onset time."

Initial position is often measured in syllables in which a vowel immediately follows the stop, but VOT can be measured in consonant clusters as well. In English, for example, liquids in initial clusters are frequently devoiced or partially devoiced (that is, aspirated) following voiceless stops (Klatt, 1975).

Some languages, including English, have voicing contrasts in word-final position. To the extent that laryngeal timing can account for such contrasts, the term *voice onset* is awkward. Here it would be voice offset time, which could perhaps be abbreviated as VOFT. Many tokens of final stops are not released in English (Davidson, 2011); in some languages, such as Thai, final stops are never released (Abramson & Tingsabadh, 1999); there may be some cases where final stops must be released, as suggested for Eastern Armenian (Henton, Ladefoged, & Maddieson, 1992, p. 86). Even if unreleased, other cues may sustain a voicing distinction, such as the duration of the preceding vowel in English (Raphael, 1972). However, the stops may still be distinguishable even without such vowel duration differences (e.g. Nittrouer, 2004).

The concept of VOT has also been applied to fricatives and affricates (Abramson, 1995) in two dialects of the Karen language (see also Jacques, 2011; Salgado, Slavic, & Zhao, 2013). The Sgaw dialect, spoken in Taunggyi, Myanmar, aside from a three-way distinction of the stops, has a distinction between unaspirated /s/ and aspirated /s^h/. In the latter, the span of local turbulence of the fricative consonant opens into normal aspiration (see first author's productions in Fig. 5). Korean plain /s/ has also been shown to be aspirated in initial position (Cho, et al., 2002, p. 212). Another example is the glottal fricative [h], which is, in initial position, essentially a vowel or diphthong excited for part of its length by turbulence through a glottal opening yielding a voicing lag. In some languages, such as certain varieties of Arabic, there is a post-vocalic [h], which is an early voicing offset with turbulence for the rest of the vowel. It is not customary to include /h/ in VOT measures, although for such varieties of Arabic laryngeal timing of postvocalic /h/ is the feature that distinguishes the normal vowel from a postaspirated one. For affricates, the closure can be voiced, but a voicing distinction can also be made without it. In English, for example, the affricates of *jest* and *chest* can be seen as lying on the VOT dimension. The distinction between frication and aspiration is not always easy to delimit, especially as the amplitude of the frication decreases. However, for aspirated fricatives and affricates, it must be apparent that such a transition has occurred. Clearly then, VOT is relevant for more phonetic classes than just stops.

We propose the following modifications (most of which have been part of the phonetics literature for some time) to the original definition:

In medial position, the duration of the voicing gap itself is usually apparent, unlike the case of stops in absolute initial position. It would be possible, then, to shift toward a more complete measure such as "VG" for Voicing Gap. However, then there would be nothing to compare with the /p/ of "pig" by itself and in the phrase "a pig." The former would have a VOT, while the latter would have a VG. The modifications instead should be:

Intervocalic stops have a negative VOT equal to the closure duration if the closure is voiced for at least half of its duration; otherwise, they have a positive VOT. This is a heuristic based on our experience; further testing is necessary to determine how often this criterion is useful. That is, the proportion of the closure that should be voiced may need to be adjusted for different languages and/or due to further research results.

Fricative VOT, in the general case, can be treated as the duration from the offset of the noise to the onset of voicing in the following vocalic segment. In English, this will often be extremely short. If voicing occurs during the frication, there is a negative VOT. That is, the noise portion generated by the constriction is equivalent to the closure portion of the stop. If voicing is present for only at the beginning of the noise, we expect that the 50% criterion used for stop prevoicing would apply. If voicing is present only at the end of the noise, the negative VOT should consist just of that duration. For the more unusual case of the aspirated fricatives, our previous example of Sgaw Karen can then be seen clearly as a distinction between a short and long VOT for fricatives. This is presumably unusual because the airflow for a fricative noise tends to diminish in amplitude at the release, while the aspiration requires that it be relatively strong.

Affricate VOT similarly treats frication as (essentially) part of the closure, so that the end of the frication counts as the release. Positive VOT will then have aspiration on the formants following. Negative VOTs will have voicing during the frication and closure. It is not clear if there are more complex cases that will need to be addressed. It does seem likely that there is greater airflow through the noise after the pressure buildup during the stop closure, which allows the aspiration to appear more easily than for a fricative alone, given that voicing distinctions driven by aspiration are common in affricates.

VOT does not seem to be as useful for stops in final position, particularly due to the commonness of unreleased stops in absolute final position. Similar measures could still be made and labeled VOT, or perhaps as VOFT to signal the syllable position.

Preaspirates and voiced aspirates should be treated outside of VOT; they have special considerations that do not fit into this limited metric.

3. Complications

There are three complications that should be mentioned. One is the effect that place of articulation can have on VOT values. Another is the matter of what zones along the timing dimension are chosen by languages. The third is the compromises entailed by ignoring the overlap of voicing and aspiration.

Using VOT data from 18 languages, Cho and Ladefoged (1999) make an important contribution to the VOT literature by focusing their attention on the increasing values of voicing lag as the place of stop articulation moves from the lips to the back of the mouth, a widely reported phenomenon (e.g., Cooper, 1991; Nearey & Rochet, 1994; Weismer, 1979). Cho and Ladefoged discuss such possible causative factors as cavity size behind the occlusion and in front of it, the size of the area of contact, and the aerodynamics in the region of the larynx. Thus, language-specific phonological rules assign VOT targets, e.g., long lag for a voiceless aspirated stop, but “automatic physiological and aerodynamic processes” as “universal phonetic implementation rules” for the amount of voicing lag (p. 236). The means of implementing such differences remain to be worked out.

Despite the universal aspect, individual differences occur in VOT. For example, Theodore, Miller and DeSteno (2009) found that there were residual differences for individuals in VOT

even after accounting for differences in speaking rate. Recent work has found further structure with this pattern, namely, that individual speakers who have longer VOTs for one category tend to have longer VOTs for the others (Chodroff & Wilson, 2017). That is, a speaker with a fairly long VOT for /k/ typically had a relatively long VOT for /t/ and /p/ as well. It seems that speakers adopt settings for the glottal gesture that are consistent across stops, resulting in the covariation. Further research has shown that this pattern extends to languages as a whole (Chodroff, Golden, & Wilson, in preparation)

The second complication has to do with the phonological choice by languages of particular zones on the VOT dimension for the distinction of consonantal categories (Cho & Ladefoged, 1999; Raphael, et al., 1995). Normally for consonantal distinctions we have three available zones on the VOT dimension: lead, short lag, and long lag. Some languages, e.g., Thai, use all three. By and large, two-category languages choose two adjacent zones out of the three. In utterance-initial position widespread varieties of English have /bdg/ with short lag and /ptk/ with long lag. In their study of Modern Hebrew (Raphael, et al., 1995), the authors find many speakers using two non-adjacent zones, voicing lead and moderately long “intermediate” values of voicing lag. They cite some other languages and discuss the factors that might be involved. Cho and Ladefoged (1999) argue that some languages choose non-canonical values of VOT, somewhat independently of the three-way distinction originally proposed. Additionally, languages can change across time, as with the Germanic language Dutch adopting an atypical (for the family) prevoiced/inaspirate distinction, or the evolution of even more elaborate patterns, as in Western Andalusian Spanish (Torreira, 2012). Further work remains to be done, both to ensure that sufficient data have been gathered and to find meaningful explanations of exceptional patterns that are found.

The third main complication is that VOT requires a separation of voicing and aspiration along the time domain, even though they frequently overlap (as in the “SA” parameter of Mikuteit & Reetz, 2007). In Fig. 6, a clearly periodic portion of the waveform can also have noticeable aspiration, as seen in the first two periods of the highlighted portion. The next three periods in the highlighted portion show decreasing amounts of simultaneous aspiration. Such overlap leads to difficulty in measuring VOT because each segment of the waveform is meant to belong only to the voiced or to the voiceless part of the signal. What is probably happening is that as the glottis approaches closure, the vocal folds begin to vibrate while some turbulence, likely to be inaudible, continues to come through the remaining shrinking glottal opening. Such turbulence ceases when the folds are fully in normal phonatory position. Although this is a difficulty, it is a familiar one in phonetics. There is often reference to the duration of “the stop” as being only closure plus voiceless portion of the signal, even though the formant transitions are equally part of the stop realization. Indeed, both voiced and voiceless transitions belong to both the consonant and the vowel, so, in one sense, one could call the aspiration a voiceless vowel. However, the voicing distinction is phonologically easier to state when the consonant governs VOT. Nonetheless, such idealizations are of great use as long as they are applied consistently. Thus, whether the region of overlap is assigned to the VOT (that is, idealized as voiceless) or to the vocalic segment (idealized as voiced), that decision should be applied to all measures within a sample, and the basis for the decision made explicit.

4. Recommendations for labeling

Measurement of VOT nowadays typically relies on both the waveform and the spectrogram. Displays such as those from Praat (Boersma, 2001) (see Fig. 7) allow the two signals to be easily compared, with temporal markings showing up in both simultaneously. Here, we will describe the labels as elements in a Praat TextGrid, though labels from any annotation program would function equivalently. Although the label names are arbitrary, adopting a standard can help with reanalysis of archival material (Simons, 2009). There is a procedure, "prepopulate.praat," that creates a new annotation tier ("Phone") that will have all the recommended VOT labels inserted for every stop. The durations are initially set to an arbitrary value of ¼ of the segment's duration. We have found that adjusting the boundaries of the existing labels is faster than entering them by hand. If a particular segment lacks any of the phonetic elements (VDCLO, VLCOS, or ASP), those can be deleted. There must be a REL label, even if it is exceedingly short, for the VOT procedure to work. The Praat procedure "get_vot" can then generate another annotation tier (defaulted to be named "VOT") that has the computed values (J. Kang & Whalen, 2017). If only some of the phonetic element labels are used for any particular dataset, VOT values can still be easily calculated. These two procedures, along with sample .wav and .TextGrid files, are available at www.github.com/HaskinsLabs/get_vot. The procedures are designed for stops; labels for affricates and fricatives would presumably include a label for the frication, which could be incorporated with some modification of the procedure(s).

It is worth remembering that the time resolutions for the waveform and the spectrogram are different. Waveforms can be realistically measured to the sample, so, for the CD rate of 44.1 kHz, to .023 ms. The nature of the signal, not to mention human measurement error, greatly increases the interval that we might want to rely on, but measurements made to the millisecond are not unrealistic. For spectra, on the other hand, there is a great deal of smearing in the temporal dimension. The default settings in Praat, for example, have a 5 ms window length, making measurements of less than 5 ms suspect. If a typical error range of +/− 1 frame is included, 15 ms is the smallest reliable individual measurement. But even those frames typically include analysis of at least 10 ms of the signal, depending on the settings of the analysis parameters. Limits of perceptibility also suggest that measurements less than 5 ms are suspect. In cases in which smaller differences appear to be statistically significant, it may be that the effect sizes will instead be small. Further study is needed in this regard.

We recommend that the events in the VOT domain be marked on an Interval Tier within Praat, or the equivalent in other programs. Intervals cannot be of zero duration, so if a component is absent from the signal, the interval must be as well. These intervals (preceding vowel (V1), voiced closure (VDCLO), voiceless closure (VLCLO), release (REL), aspiration (ASP) and following vowel (V2)) encompass the acoustic segments used in our definition of VOT (see Fig. 7):

(V1) VDCLO VLCLO REL ASP (V2)

For those stops with both voiced and voiceless closure, a decision has to be made about whether the VOT is positive or negative (see above). In the current version of the get_vot,

the stop is designated as prevoiced if the voiced closure is at least half the duration of the closure. This is an assumption that can be addressed by individual researchers. It is easy to change this parameter in the procedure by changing the “pct_voicing” variable at the beginning of the procedure to be whatever percentage of the closure duration the researcher finds to be appropriate for the measurements’ use. Indeed, the issue has not been fully addressed in the literature.

The vowel durations (V1 preceding, V2 following) are not strictly part of the VOT measurement but are often important for intuiting the perceptual value of the other measurements. VDCLO begins at the point of closure if there is voicing present, and ends either when the voicing ceases or at release; this can either be “edge vibrations” (Davidson’s “bleed”) or full voicing. Note that closure is typically apparent in a large decrease in intensity along with decrease or elimination of the higher formants. Occasionally, a “closure burst” of very short duration can be seen. Some instances of closure can be difficult to locate with certainty. VLCLO begins at closure (if there is no VDCLO) or else at the end of VDCLO, and it ends at release. REL begins at the onset of the burst (if present) or the onset of the vocalic formants, and ends either when the aspiration begins or, if there is none, at the end of the noise of the burst. (If there is essentially no burst and no aspiration, this current scheme requires that an extremely short REL be marked to anchor the later measurements; if, on the other hand, the phonetic segment is lenited to the extent that it is no longer a stop, it should not be given a VOT measurement.) ASP begins at the end of REL and ends when the noise gives way to the voiced vocalic segment. Note that there is often overlap between noise and voicing, giving rise to differences in judgment about where the crossover should be (see Fig. 6); this is an unavoidable consequence of putting a sharp division between two otherwise distinct objects just where one gives way to the other, much as “night” and “day” are extremely different but difficult to divide at an instant in time during twilight.

We should note that we have treated the release as one acoustic event even though some authors divide it into a “transient” followed by “frication” (e.g., Stevens, 1993). Although this distinction has some grounding in the physics of the aerodynamics at release, the practical matter of distinguishing them has proven to be quite difficult and unreplicable. (Distinguishing the end of the burst from the onset of the aspiration is challenging enough.) For VOT, the separation of transient and frication holds no theoretical importance, so, the two together are classified as the release.

5. Automatic measures

In recent years, several systems have been proposed to measure VOT automatically (e.g., Das & Hansen, 2004; Hansen, Gray, & Kim, 2010; Kazemzadeh, et al., 2006; Lin & Wang, 2011; Sonderegger & Keshet, 2012; Stouten & Van hamme, 2009). Building on the greatly improved success of speech segmentation in general, these programs have been reported to have a reasonable degree of success. Some work only on initial stops while others label medial stops as well. The most widely used system (Keshet, Sonderegger, & Knowles, 2014) has about 80% agreement between manual measurements and the automatic ones within 5 ms, and closer to 90% at 10 ms. This system has been used for an extensive analysis of Scottish English (Stuart-Smith, Sonderegger, Rathcke, & Macdonald, 2015), where they

used a manual check on the VOTs before final analysis. They found that 62.6% of the VOT measurements were correct, 15.8% needed to be corrected by hand, and 21.6% were not usable (p. 518). The correction time was found to be much shorter than initial measurements by hand. It can be expected that these systems will continue to improve in the coming years.

6. Summary

Voice Onset Time (VOT) has proven to be a robust measure of the acoustic realization of consonantal voicing distinctions in most languages. It does not cover every distinction related to laryngeal timing and consonant articulation, nor was it ever claimed to do so. Beginning from its primary definition for stops in absolute initial position, it can be extended to intervocalic and even final position, with appropriate changes in terminology. It can also be extended to cover affricates and the (rare) aspirated fricatives. When we remember that VOT is an assignment of strict boundaries to physiological events that overlap, we can see that discrepancies in measurement are to be expected, though they can be mitigated. The full range of articulatory and acoustic aspects of the devoicing gesture underlying stop consonant distinctions remains to be fully elucidated. Nonetheless, the hundreds of studies that have used VOT in the past 50 years are a strong testament to the lasting value of this measure.

Acknowledgments

The writing of this paper was supported by NIH grant DC-002717 to Haskins Laboratories. We thank Mark K. Tiede, Laura L. Koenig, Stephanie Kakadelis, Hosung Nam, Jaekoo Kang, Lisa Davidson, Taehong Cho and two anonymous reviewers for helpful comments, and Hannah M. King for help with the figures.

References

Abramson AS. Laryngeal timing in consonant distinctions. Phonetica. 1977; 34:295–303. [PubMed: 594164]

Abramson, AS. Laryngeal timing in Karen obstruents. In: Bell-Berti, F., Raphael, L., editors. Producing speech: Contemporary issues. For Katherine Safford Harris. Woodbury, NY: American Institute of Physics Press; 1995. p. 155-165.

Abramson AS, Lisker L. Voice onset time in stop consonants: Acoustic analysis and synthesis. Rapports du 5e Congrès International d'Acoustique. 1965; 1a:A51.

Abramson, AS., Lisker, L. Laryngeal behavior, the speech signal, and phonological simplicity. In: Graur, A., editor. Actes du Xe Congrès International des Linguistes 1967. Bucarest: Éditions de l'Académie de la République Socialiste de Roumanie; 1969. p. 123-129.

Abramson, AS., Lisker, L. Voice timing in Korean stops. In: Rigault, A., Charbonneau, R., editors. Proceedings of the Seventh International Congress of Phonetic Sciences, 1971. The Hague: Mouton; 1972. p. 439-446.

Abramson, AS., Lisker, L. Relative power of cues: Fo shift versus voice timing. In: Fromkin, VA., editor. Phonetic linguistics: Essays in honor of Peter Ladefoged. New York: Academic Press; 1985. p. 25-33.

Abramson AS, Tingsabadh K. Thai final stops: Cross-language perception. Phonetica. 1999; 56:111–122.

Baran JA, Laufer MZ, Daniloff RG. Phonological contrastivity in conversation: A comparative study of voice onset time. Journal of Phonetics. 1977; 54:339–350.

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

Bennett R. Contrast and laryngeal states in Tz'utujil. The UCSC Linguistics Research Center. 2010:93–120.

Boersma P. Praat, a system for doing phonetics by computer. Glot International. 2001; 5:341–345.

Browman CP, Goldstein LM. Articulatory phonology: An overview. Phonetica. 1992; 49:155–180. [PubMed: 1488456]

Cho T, Jun S-A, Ladefoged P. Acoustic and aerodynamic correlates of Korean stops and fricatives. Journal of Phonetics. 2002; 30:193–228.

Cho T, Keating PA. Articulatory and acoustic studies on domain-initial strengthening in Korean. Journal of Phonetics. 2001; 29:155–190.

Cho T, Ladefoged P. Variation and universals in VOT: Evidence from 18 languages. Journal of Phonetics. 1999; 27:207–229.

Cho T, Lee Y, Kim S. Prosodic strengthening on the/s/-stop cluster and the phonetic implementation of an allophonic rule in English. Journal of Phonetics. 2014; 46:128–146.

Cho T, McQueen JM. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. Journal of Phonetics. 2005; 33:121–157.

Chodroff E, Golden A, Wilson C. [[Covariation of voice onset time: a universal aspect of phonetic realization]]. (in preparation).

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

Cooper, AM. An articulatory account of aspiration in English. Unpublished Ph.D. dissertation. Yale University: 1991.

Crystal TH, House AS. Segmental durations in connected speech signals: Preliminary results. Journal of the Acoustical Society of America. 1982; 72:705–716. [PubMed: 7130529]

Das S, Hansen JHL. Detection of voice onset time (VOT) for unvoiced stops (/p/,/t/,/k/) using the Teager energy operator (TEO) for automatic detection of accented English. Proceedings of the 6th Nordic Signal Processing Symposium-NORSIG. 2004; 2004:344–347.

Davidson L. Characteristics of stop releases in American English sponta neous speech. Speech Communication. 2011; 53:1042–1058.

Davidson L. Variability in the implementation of voicing in American English obstruents. Journal of Phonetics. 2016; 54:35–50.

Davis K. Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. Journal of Child Language. 1995; 22:275–305. [PubMed: 8550724]

Dixit, RP. Neuromuscular aspects of laryngeal control: With special reference to Hindi. Unpublished Ph.D. dissertation . University of Texas, Austin: 1975.

Dixit, RP. Inadequacies in phonetic specifications of some laryngeal features: evidence from Hindi. In: Hollein, H., Hollein, P., editors. Current issues in the phonetic sciences. Amsterdam: John Benjamins; 1979. p. 423-434.

Dmitrieva O, Llanos F, Shultz AA, Francis AL. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. Journal of Phonetics. 2015; 49:77–95.

Docherty, GJ. The timing of voicing in British English obstruents. Berlin/New York: Foris; 1992.

Dutta, I. Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt. Unpublished Ph.D. dissertation. University of Illinois at Urbana-Champaign: 2007.

Flemming, ES. Auditory representations in phonology. New York: Routledge; 2013.

Gallagher G. Acoustic and articulatory features in phonology - the case for [long VOT]. The Linguistic Review. 2011; 28:281–313.

Gallagher G. Natural classes in cooccurrence constraints. Lingua. 2015; 166(Part A):80–98.

Goldstein LM, Browman CP. Representation of voicing contrasts using articulatory gestures. Journal of Phonetics. 1986; 14:339–342.

Haggard M, Ambler S, Callow M. Pitch as a voicing cue. Journal of the Acoustical Society of America. 1970; 47:613–617. [PubMed: 5439661]

Hansen JHL, Gray SS, Kim W. Automatic voice onset time detection for unvoiced stops (/p/,/t/,/k/) with application to accent classification. Speech Communication. 2010; 52:777–789.

Helgason, P. Preaspiration in the Nordic languages: Synchronic and diachronic aspects. Unpublished Ph.D. dissertation. Stockholm University: 2002.

Henton CG, Ladefoged P, Maddieson I. Stops in the world's languages. Phonetica. 1992; 49:65–101. [PubMed: 1615036]

Hombert J-M, Ohala JJ, Ewan WG. Phonetic explanations for the development of tones. Language. 1979; 55:37–58.

Jacques G. A panchronic study of aspirated fricatives, with new evidence from Pumi. Lingua. 2011; 121:1518–1538.

Kagaya R. A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics. 1974; 2:161–180.

Kang, J., Whalen, DH. get_vot. 2017. https://github.com/HaskinsLabs/get_vot

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

Kazemzadeh, A., Tepperman, J., Silva, JF., You, H., Lee, S., Alwan, AA., Narayanan, S. INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing. Pittsburgh, PA: 2006. Automatic detection of voice onset time contrasts for use in pronunciation assessment; p. 721-724.

Keating PA. Phonetic and phonological representation of stop consonant voicing. Language. 1984; 60:286–319.

Keshet, J., Sonderegger, M., Knowles, T. AutoVOT, v 0.93. 2014. https://github.com/mlml/autovot/

Kim C-W. On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). Word. 1965; 21:339–359.

Kim C-W. A theory of aspiration. Phonetica. 1970; 21:107–116.

Kim M-R, Beddor PS, Horrocks J. The contribution of consonantal and vocalic information to the perception of Korean initial stops. Journal of Phonetics. 2002; 30:77–100.

Klatt DH. Voice onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research. 1975; 18:686–706. [PubMed: 1207100]

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

Ladefoged P, Ladefoged J, Turk A, Hind K, Skilton SJ. Phonetic structures of Scottish Gaelic. Journal of the International Phonetic Association. 1998; 28:1–41.

Lee HB. Korean. Journal of the International Phonetic Association. 1993; 23:28–31.

Liberman AM, Delattre PC, Cooper FS. Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech. 1958; 1:153–167.

Lin C-Y, Wang H-C. Automatic estimation of voice onset time for word-initial stops by applying random forest to onset detection. Journal of the Acoustical Society of America. 2011; 130:514–525. [PubMed: 21786917]

Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. Word. 1964; 20:384–422.

Lisker L, Abramson AS. Some effects of context on voice onset time in English stops. Language and Speech. 1967; 10:1–28. [PubMed: 6044530]

Lisker, L., Abramson, AS. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia; 1970. The voicing dimension: Some experiments in comparative phonetics.

Lisker L, Abramson AS. Distinctive features and laryngeal control. Language. 1971; 47:767–785.

Lisker, L., Abramson, AS. Phonetic validation of distinctive features: A test case in French. In: Channon, R., Shockey, L., editors. In honor of Ilse Lehiste. Dordrecht: Foris; 1987. p. 183-190.

Lisker L, Abramson AS, Cooper FS, Schvey MH. Transillumination of the larynx in running speech. Journal of the Acoustical Society of America. 1969; 45:1544–1546. [PubMed: 5803181]

Löfqvist A, Baer T, McGarr NS, Story RS. The cricothyroid m uscle in voicing control. Journal of the Acoustical Society of America. 1989; 85:1314–1321. [PubMed: 2708673]

Martin SE. Korean phonemics. Language. 1951; 27:519–533.

Mikuteit S, Reetz H. Caught in the ACT: The timing of aspiration and voicing in East Bengali. Language and Speech. 2007; 50:247–277. [PubMed: 17702474]

Nance C, Stuart-Smith J. Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. Journal of the International Phonetic Association. 2013; 43:129–152.

Nearey TM, Rochet BL. Effects of place of articulation and vowel context on VOT production and perception in French and English stops. Journal of the International Phonetic Association. 1994; 24:1–19.

Nittrouer S. The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. Journal of the Acoustical Society of America. 2004; 115:1777–1790. [PubMed: 15101656]

Panconcelli-Calzia, G. Die experimentelle Phonetik in ihrer Anwendung auf die Sprachwissenschaft. Berlin: Walter de Gruyter; 1924.

Raphael LJ. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. Journal of the Acoustical Society of America. 1972; 51:1296–1303. [PubMed: 5032946]

Raphael, LJ., Tobin, Y., Faber, A., Most, T., Kollia, HB., Milstein, D. Intermediate values of Voice Onset Time. In: Bell-Berti, F., Raphael, LJ., editors. Producing speech: Contemporary issues. For Katherine Safford Harris. Woodbury, NY: American Institute of Physics Press; 1995. p. 117-127.

Repp BH. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. Language and Speech. 1979; 22:173–189. [PubMed: 513908]

Rousselot, P-J. Principes de phonétique expérimentale. Paris: H. Welter; 1897–1908.

Salgado H, Slavic J, Zhao Y. The production of aspirated fricatives in Sgaw Karen. Studies in the Linguistic Sciences: Illinois Working Papers. 2013; 2013:148–161.

Sawashima M, Abramson AS, Cooper FS, Lisker L. Observing laryngeal adjustments during running speech by use of a fibreoptics system. Phonetica. 1970; 22:193–201.

Shimizu, K. A cross-language study of voicing contrasts of stop consonants in Asian languages. Seibido; 1996.

Shockey, L. Sound patterns of spoken English. Malden, MA: Blackwell; 2003.

Silva DJ. Acoustic evidence for the emergence of tonal contrast in cotemporary Korean. Phonology. 2006; 23:287–308.

Silverman D. On the rarity of pre-aspirated stops. Journal of Linguistics. 2003; 39:575–598.

Simons, GF. Linguistics as a community activity: The paradox of freedom through standards. In: Lewis, WD.Karimi, S.Harley, H., Farrar, S., editors. Time and again: Theoretical perspectives on formal linguistics: In honor of D. Terence Langendoen. Amsterdam: John Benjamins; 2009. p. 235-250.

Sonderegger M, Keshet J. Automatic measurement of voice onset time using discriminative structured prediction. Journal of the Acoustical Society of America. 2012; 132:3965–3979. [PubMed: 23231126]

Steriade, D. Paradigm uniformity and the phonetics-phonology boundary. In: Broe, MB., Pierrehumbert, JB., editors. Papers in laboratory phonology V: Acquisition and the lexicon. Cambridge: Cambridge University Press; 2000. p. 313-334.

Stevens KN. Models for the production and acoustics of stop consonants. Speech Communication. 1993; 13:367–375.

Stouten V, Van hamme H. Automatic voice onset time estimation from reassignment spectra. Speech Communication. 2009; 51:1194–1205.

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

Theodore RM, Miller JL, DeSteno D. Individual talker differences in voice-onset-time: Contextual influences. Journal of the Acoustical Society of America. 2009; 125:3974–3982. [PubMed: 19507979]

Tillmann, HG. Early modern instrumental phonetics. In: Koerner, EFK., Asher, RE., editors. Concise history of the language sciences: From the Sumerians to the cognitivists. Oxford: Pergamon; 1995. p. 401-416.

Torreira F. Investigating the nature of aspirated stops in Western Andalusian Spanish. Journal of the International Phonetic Association. 2012; 42:49–63.

Weismer G. Sensitivity of voice onset measures to certain segmental features in speech production. Journal of Phonetics. 1979; 7:194–204.

Westbury JR. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. Journal of the Acoustical Society of America. 1983; 73:1322–1336. [PubMed: 6853844]

Wetzels WL, Mascaró J. The typology of voicing and devoicing. Language. 2001; 77:207–244.

Whalen DH, Abramson AS, Lisker L, Mody M. Gradient effects of fundamental frequency on stop consonant voicing judgments. Phonetica. 1990; 47:36–49. [PubMed: 2277812]

Whalen DH, Abramson AS, Lisker L, Mody M. F0 gives voicing information even with unambiguous voice onset times. Journal of the Acoustical Society of America. 1993; 93:2152–2159. [PubMed: 8473630]

Highlights

- We review the 50 year history of the concept of Voice Onset Time (VOT)
- This widely-used acoustic measure of consonantal voicing distinctions retains its usefulness.
- Complications are discussed, and a standard set of labels proposed.

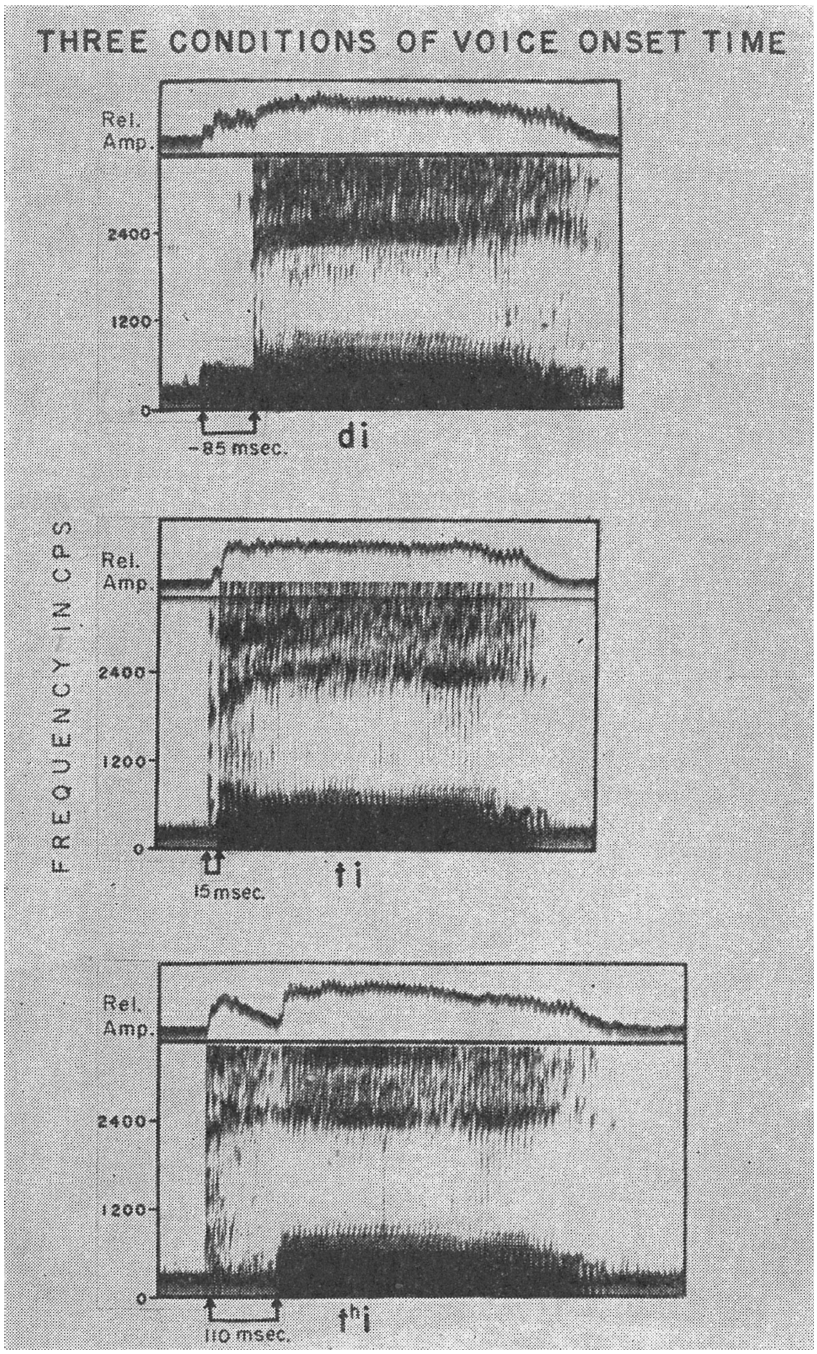


Fig. 1.

Examples of the three main categories of stop (in this case, for Thai). From Lisker & Abramson (1964); used by permission.

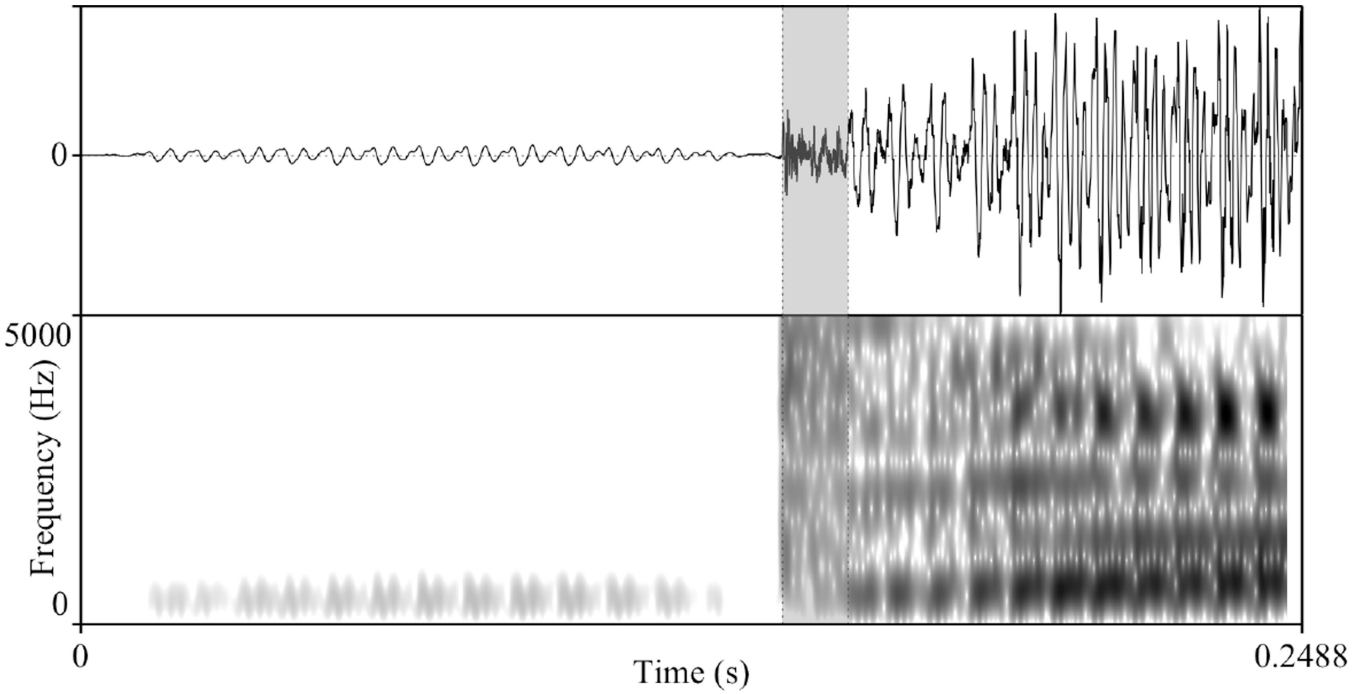


Fig. 2.
Praat display of the beginning portion of an English prevoiced [ga]. The burst is highlighted in grey.

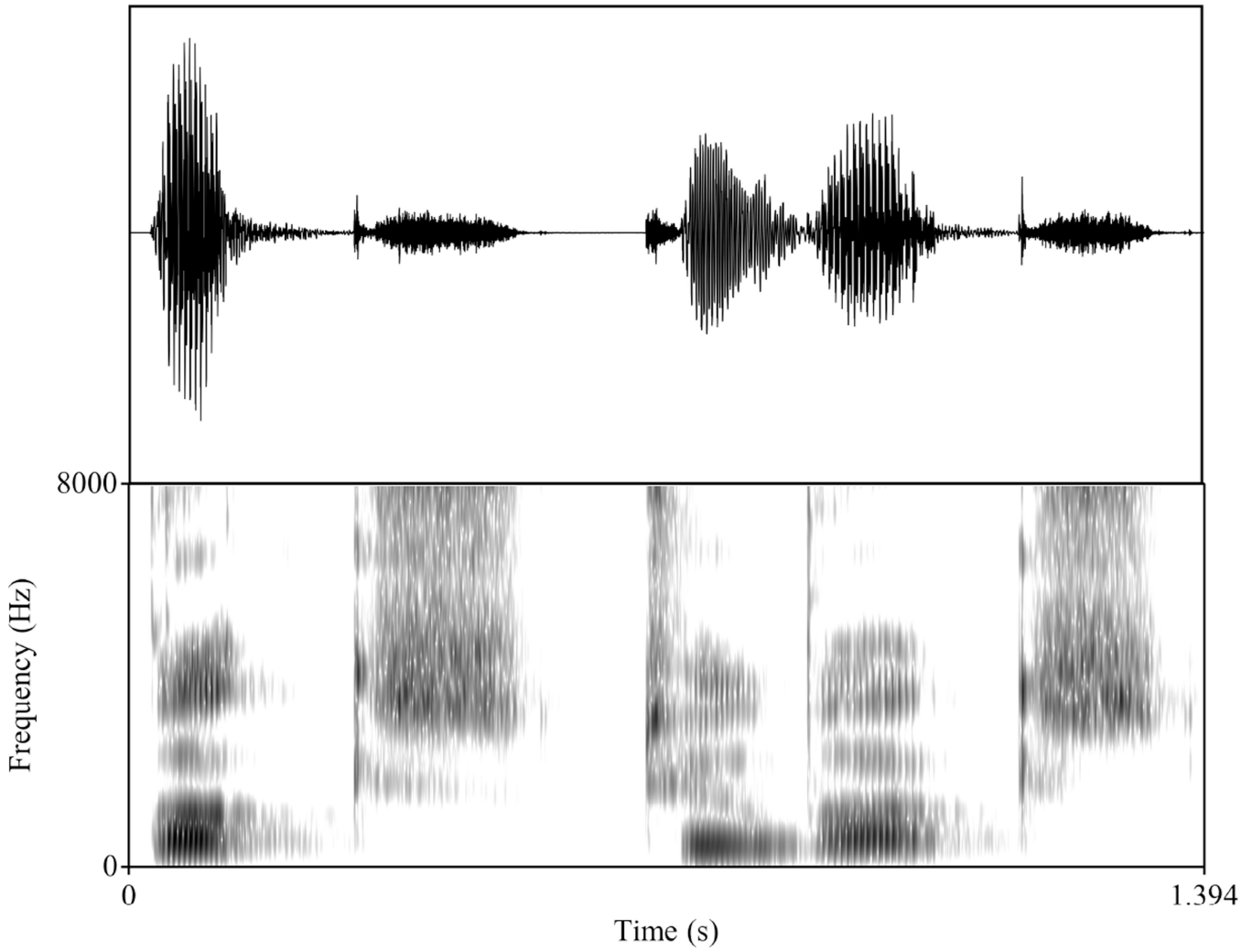


Fig. 3.

Production of “books” (left) and “two books” right by the second author. The first [b] has a 5 ms VOT, while the second has a −67 ms VOT (with 11 ms of voicelessness following the release).

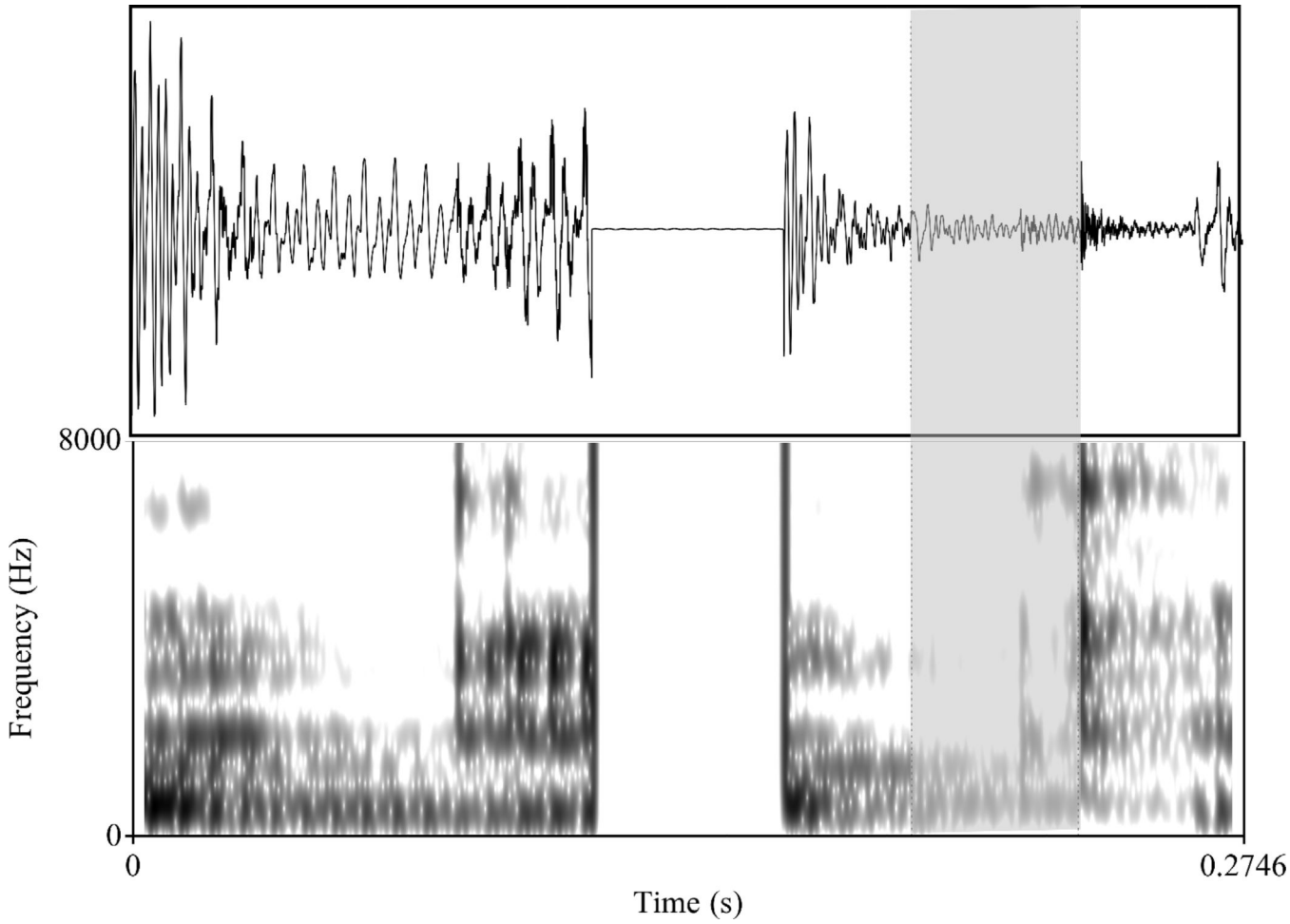


Fig. 4.

Examples of closures from “tugging” (left) and “tucking” right as produced by the second author. The voiceless “closure” of the [k] is highlighted in grey; the noise indicates incomplete closure.

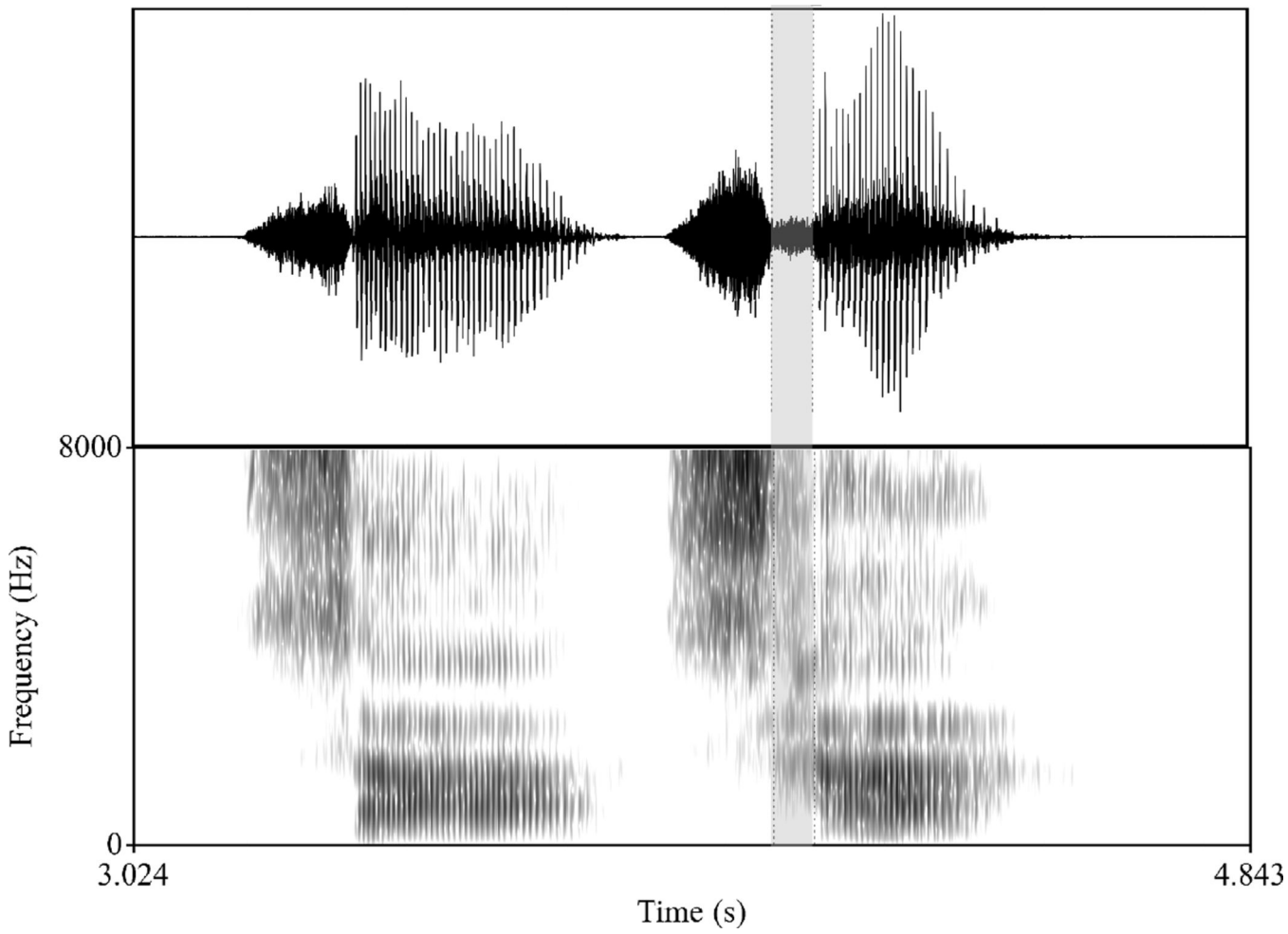


Fig. 5.

Example of aspirated [s], from the first author's production. The first syllable is [sa] with a 0 ms VOT; the second is [s^ha], with the positive VOT (75 ms) marked in grey shading.

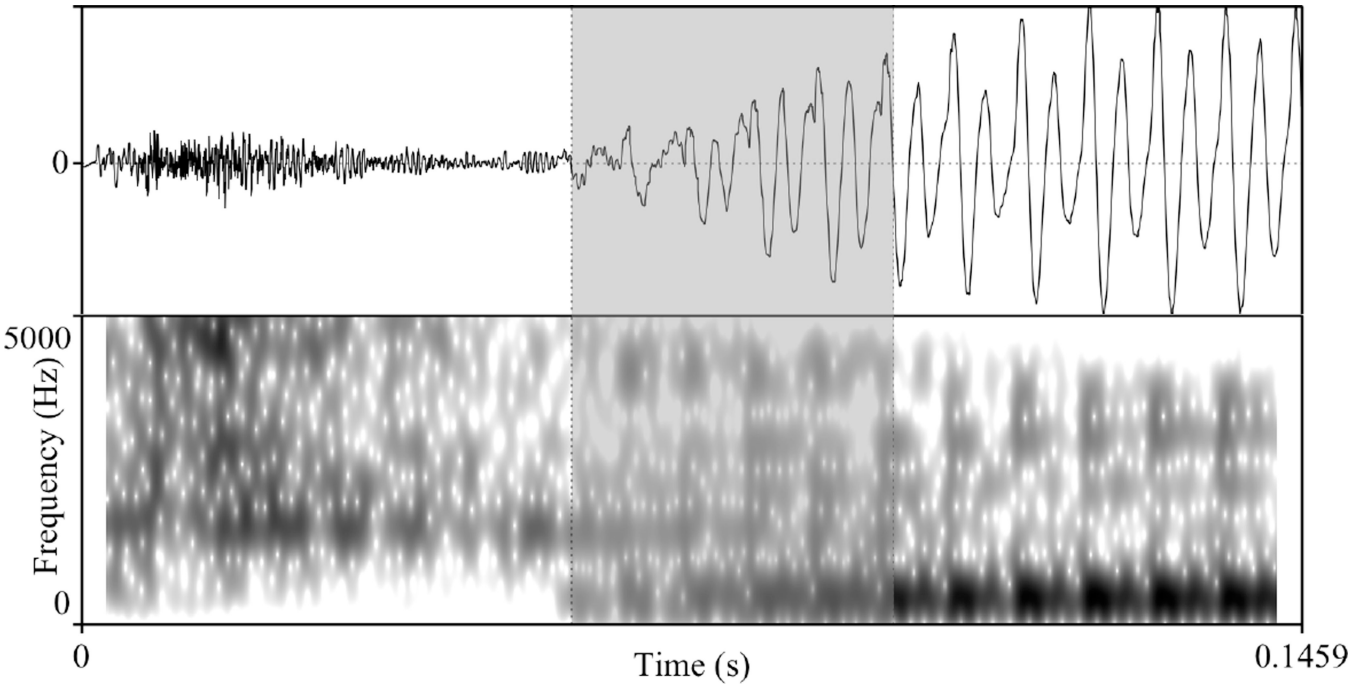
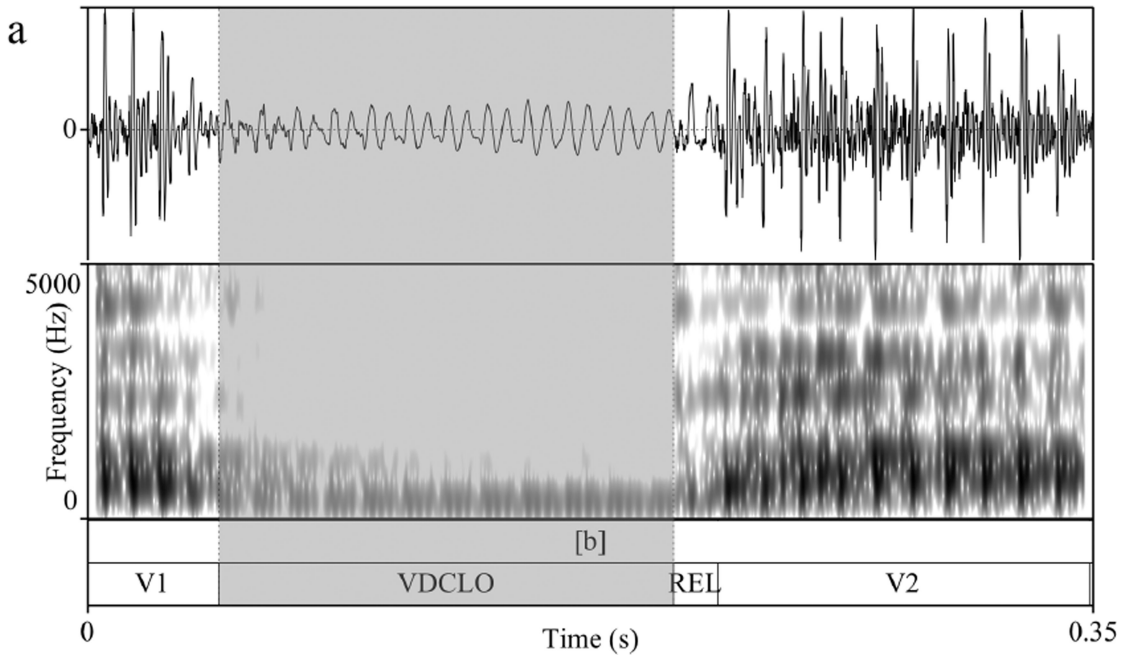


Fig. 6.

Overlap of voicing and aspiration. Selection of the utterance [$k^h\alpha$] from a speaker of American English. Highlighted region shows 2–5 glottal pulses of overlap with aspiration.



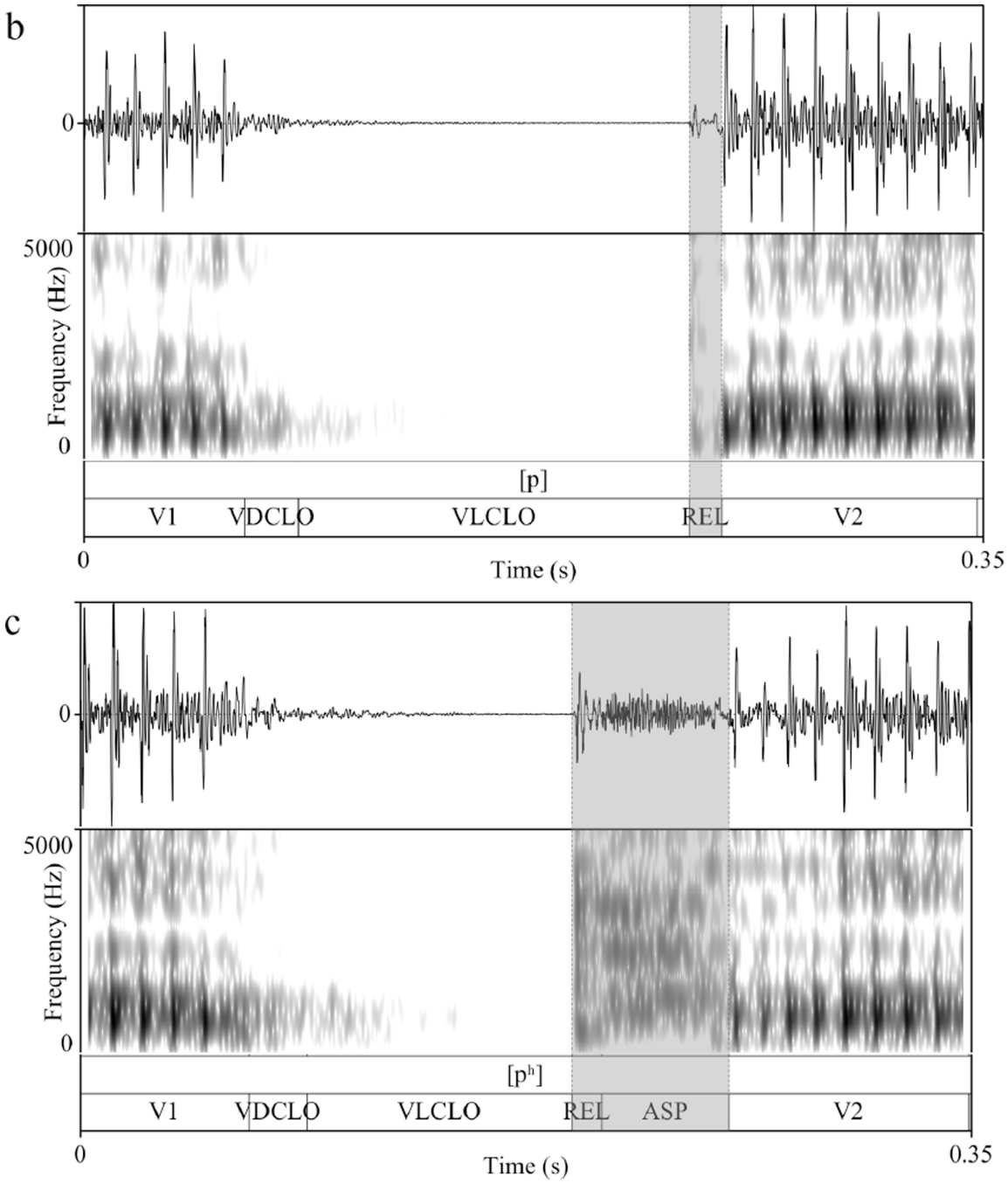


Fig. 7.

Examples of VOTs in [ɑCɑ], with the consonant varying in each panel. a) [b], b) [p], and c) [p^h]. Produced by the first author, a speaker of English, bilingual in Thai. The VOTs are highlighted in grey, with the values: a) −158 ms, b) 13 ms, and c) 62 ms.