Voice onset time in aphasia, apraxia of speech and dysarthria: a review

PASCAL AUZOU†, CANAN ÖZSANCAK‡, RICHARD J. MORRIS§, MARY JAN‡, FRANCIS EUSTACHE¶ and DIDIER HANNEQUIN‡

†Fédération de Neurologie (Rouen, F) et Unité INSERM 320 (Caen, F)
‡Fédération de Neurologie (Rouen, F)
§Department of Communication Disorders, Tallahassee, Florida 32306-1200, USA
¶Service de neurologie et Unité INSERM 320 (Caen, F)

(Received 14 August 1998; accepted 7 June 1999)

Abstract

Voice onset time (VOT) is an objective temporal acoustic parameter defined as the time between the release of the oral constriction for plosive production and the onset of vocal fold vibrations. Many researchers consider VOT to be the most reliable acoustic cue for the distinction between voiced and voiceless stops. Previous studies have explored the physiological and linguistic factors underlying VOT production in normal speakers across several languages. A major clinical goal of acoustic analysis in speech disorder is to establish a correlation between the acoustic abnormalities and the phonetic perturbations. VOT could thus be used as an acoustic parameter that indicates the phonetic contrast between voiced and voiceless stops. This paper includes a critical review of the measurement of VOT, factors of VOT variability and the effect of neurogenic communication disorders on VOT. We review the VOT data from subjects who exhibit aphasia, apraxia of speech and dysarthria. These studies reveal that VOT perturbations in aphasia have been interpreted as phonemic or phonetic errors, while VOT abnormalities in apraxia of speech and dysarthria grossly reflect loss of motor control.

Keywords: voice onset time, acoustic, aphasia, apraxia of speech, dysarthria.

Introduction

Voice onset time (VOT) is an objective acoustic parameter reflecting motor speech control. It has been measured in normal speakers and those with speech disorders to test linguistic hypotheses or to measure the timing between the structures of the
larynx and the oral cavity. VOT is defined as the time between the release of the oral constriction for plosive production and the onset of vocal-fold vibration (Lisker and Abramson, 1964; 1967; Zlatin, 1974). This temporal parameter is the most reliable acoustic cue for the distinction between voiced and voiceless stops. These consonants are characterized by the creation of a pressure pulse during the complete occlusion in the vocal tract.

Production of word-initial voiceless stops involves a period of articulatory closure during which the vocal folds are maintained in a relatively open position without glottal pulsing. During the closure phase, intraoral pressure increases rapidly until it becomes equal to the subglottal pressure (Hertegard and Guaffin, 1995). The equalization of the two pressures creates an aerodynamic environment that inhibits vocal fold vibration. The vocal folds are also actively abducted to interrupt voicing. To end the closure interval, the oral contact is released, resulting in a rapid flow of air out of the mouth and a rapid drop in the intraoral air pressure. This airflow creates a transient burst of acoustic energy. Although the glottis remains open at the moment of articulatory release, the vocal folds then approximate to begin the glottal vibration of the following vowel.

The duration of the abduction-adduction gesture of the vocal folds is a primary cue to the phonetic voiceless–voiced distinction between the plosive cognates (Benguere1, Hirose, Sawashima and Ushijima, 1978; Hirose and Gay, 1972). Abduction of the vocal folds is mainly controlled by the posterior cricoarytenoid muscle and their adduction is mainly controlled by the interarytenoid muscle (Cooper, 1992). Hence, the accurate production of stop consonants requires close coordination between the larynx and the lips, tongue and jaw. It has been suggested that the glottal opening gesture, maximum glottal opening, and articulatory closure release may be coordinated in terms of a fixed temporal ‘target’ (Zebrowski, Conture and Cudahy, 1985). The timing control of related articulators to achieve a temporal target suggests that the control of speech movements occurs over aggregates rather than individual articulators (Löfqvist and Lindblom, 1994; Munhall, Löfqvist and Kelso, 1994). These speech movements might be comprised of a set of motor programs specified for the individual phonemes that would be activated and sequenced into larger aggregates associated with syllables, words and phrases to allow meaningful communication (Gracco and Löfqvist, 1994).

In nearly all languages VOT plays a primary role in distinguishing between voiced and voiceless plosive cognates (Lisker and Abramson, 1964). However, a complete acoustical description of plosive consonants must take into consideration other parameters such as the presence or absence of fricative noise upon consonantal release, intensity of the noise and duration of the burst release or formant transitions (De Mori and Flammia, 1993; Lisker and Abramson, 1964; Stevens and Klatt, 1974; Sussman and Shore, 1996). VOT denotes the coarticulatory timing control between laryngeal and oral articulation. Factors that can affect the magnitude of VOT include size of glottal opening, transglottal pressure and vocal fold tension (Löfqvist, 1992; Tyler and Watterson, 1991).

The purpose of this paper is to present a review of current information on VOT. The first section includes a review of the methodological considerations when making VOT measurement. Then, a section on normative VOT data follows. After that is a section on the linguistic and physiological variations that can alter VOT values in normal speech production. The final section includes a review of the information on VOT production in subjects who have aphasia, apraxia of speech and dysarthria.
The measurement of VOT: methodological considerations

The measurement of VOT is usually obtained from wideband spectrograms according to the procedure first recommended by Lisker and Abramson (1964). That is, VOT corresponds to the interval between the onset of the energy 'burst', representing the release of an articulatory constriction, and the first of the regularly spaced vertical striations representing the vocal fold vibration (figure 1). The instant of release is defined as the point where the spectrogram shows an abrupt spectral change representing the transient noise burst. When the vocal fold vibration precedes the release as in voiced stops, the VOT is given a negative value and is called voicing lead (Lisker and Abramson, 1964; Macken and Baron, 1980a,b; Zlatin and Koenigsknecht, 1976). On the other hand, when the release precedes the vocal fold vibration as in voiceless stops, the VOT value is positive and called voicing lag.

Lag time measurements are often taken from the burst to the onset of the first formant (F1) of the following vowel. In some studies, lag time measurements were taken from the closure release at the beginning of the burst to the beginning of the second formant (F2) of the following vowel (Gandour, 1985; Klatt, 1975). The choice of the second formant as a delimiter of VOT is particularly functional for voiced stops because voicing frequently occurs during the lag time, making it difficult to determine exactly where the first formant of the vowel begins (Davis, 1995).

Figure 1. Oscillogram of the syllable /ka/ in a normal subject (top) and the corresponding wide-band spectrogram (bottom). VOT represents the duration of time between the opening of the lips (a on the oscillogram and a’ on the spectrogram) and the regular striations representing the vocal fold vibrations (b on the oscillogram and b’ on the spectrogram).
Generally, some delay occurs between the onset of F2 and F1. However, this difference is three milliseconds or less for both short and long lag stops. Thus, using F2 onset does not introduce an important measurement error.

VOT measurements are usually done from isolated words or syllables or from words or syllables embedded in carrier phrases (e.g. Say ___ again; Say ___ instead). Using a carrier phrase has two effects on the subject’s speaking rate. First, the longer utterance causes the subject to time the utterance phonemes and syllables in a manner similar to conversational speech rather than the prolongation common to citation speech. Second, the effects of prepausal phoneme lengthening in utterance final syllables is removed by embedding the target phoneme in the middle of the phrase.

Digitized acoustic signals allow the experimenter to obtain time-synchronized wideband spectrographic and oscillographic displays. Rather than using the spectrogram, several authors proposed directly measuring VOT from the oscillogram (Keller, 1990; Lane, Wozniak and Perkell, 1994; Tyler and Watterson, 1991; Volaitis and Miller, 1992). Using a display of the digitized speech signal, these authors measured VOT from the onset of the plosive release burst to the first zero crossing at the onset of periodicity in the waveform. However, others report that the simultaneous examination of oscillograms and spectrograms provides a greater accuracy of measurement (Bortolini, Zmarich, Fior and Bonifacio, 1995; Davis, 1995; Petrosino, Colcord, Kurcz and Yonker, 1993). Whenever precise measurement is questionable on the spectrogram, the oscillographic display can be used to substantiate onset of either the release burst or voicing for the vowel.

A physiologic variation in measuring VOT was reported by Morris and Brown (1987). They used the oscillographic displays from intraoral air pressure changes and a simultaneous signal from a contact microphone on the neck to assess VOT. Brown, Morris and Weiss (1993) reported that this system provided output comparable to measures made with spectrograms and expanded digital oscillographic displays.

In some utterances, it is not possible to determine the location of the burst (figure 2) or the onset of regular striations on the spectrogram (figure 3). For example, the burst does not occur when a subject fails to achieve full closure in the production of stop consonants. Similarly, the vowel onset may be difficult to determine (Sweeting and Baken, 1982). Lisker and Abramson (1964) noted the risk of including a short span of pulsation, as brief as a single cycle of vocal fold movement, that would be too weak to be audible (figure 4). They believed that such occasional errors make no significant difference in the use of VOT to categorize stop consonants. The combined use of oscillograms and spectrograms may reduce the frequency of such errors. In addition, a double burst may make it difficult to determine when the release the lip or tongue closure occurs (figure 5). In these cases, measurements are usually taken from the start of the first burst (Davis, 1995). Finally, few researchers report the percentage of unmeasurable VOTs; the data available indicate that less than 4% of the productions cannot be measured (Itoh, Sasanuma, Tatsumi, Murakami, Fukusako and Suzuki, 1982; Sweeting and Baken, 1982).

Several authors report that the reliability of VOT measurements is high. Connor, Ludlow and Schultz (1989) completed two analyses of the spectrograms produced by 12 normals and 12 dysarthric patients with Parkinson’s disease (PD). The inter-measurement reliability was 4.5% (range 4.1–6.8%) for the syllables produced by the normal subjects and 4.4% (range 1.7–6.7%) for the syllables produced by the PD subjects. Caruso and Burton (1987) found mean intrajudge measurement errors.
of 5.56 msec, and interjudge ones of 4.75 msec. Other authors have reported closer agreement in terms of percentage agreement within a selected range of duration differences, such as 96.6% of reliability measures being within 2 msec (Sweeting and Baken, 1982), 100% agreement within 3 msec (Davis, 1995) and 97.5% agreement within 5 msec (Zlatin, 1974). Similarly, Hoit, Solomon and Hixon (1993) reported 94.3% interjudge agreement and 95.1% intrajudge agreement.

**Norms of VOT production**

VOT data have been reported for normal speakers of many languages. These studies indicate similar VOT categories across languages. Lisler and Abramson (1964) reported mean VOT values for word-initial stop consonants data from normal subjects who spoke 11 different languages. However, each language was represented by a small pool of subjects. The greatest number of studies have been done using English speaking subjects. The results from many of these studies are shown in table 1. The VOT values vary for each consonant and may be best represented as a range of durations. Table 1 reveals that the range of VOT values produced for a given consonant is comparable across studies and across different phonetic contexts. Extensive cross-language studies by Lisler and Abramson (1964, 1967) demonstrated that three categories of stops emerge along the VOT continuum, with reasonable agreement as to category boundaries across languages.

1. 'Voicing lead': negative VOT values, ranging from about −125 to −75 msec, with a median value of −100 msec. Italian or French voiced stops are of this type (Bortolini et al., 1995; Ryalls, Provost and Arsenault, 1995).
Figure 3. Same representation as in Fig. 1 for the [ta] of a patient with spastic dysarthria. The striation does not become regular enough to determine the beginning.

(2) ‘Short voicing lag’: positive VOT values, ranging from 0 to +25 msec, with a median value of +10 msec. English voiced stops and Italian voiceless stops are of this type.

(3) ‘Long voicing lag’: large positive VOT values, ranging from +60 to +100 msec, with a median value of +75 msec. English voiceless stops are of this type.

However, Lisker and Abramson (1964) reported that in English the phonemes [b], [d] and [g] used two of the VOT categories. Their group of four speakers exhibited VOT values for [b] that ranged from either −130 to −20 msec (voicing lead) or from 0 to +5 msec (short lag). The individual speakers did not tend to exhibit both VOT categories. One speaker produced 95% of all the stops with voicing lead, and a second one produced the remaining 5%. The other two speakers produced almost exclusively positive values or short lag (41/42 for one speaker and 56/58 for the other). Because the voiced stops in English occur in two VOT categories whereas the voiceless stops occur in one VOT category, the VOT values of voiceless stops tend to be more normally distributed than their voiced counterparts (Zlatin, 1974).

Thai is a language that uses all three voicing categories. Lisker and Abramson
Figure 4. Oscillogram and spectrogram of a [ka] in a normal subject. If using only the oscillogram, the beginning of the periodic vibration may be doubtful (b). When using combined oscillogram and spectrogram, the ambiguity disappears (correct beginning in b').

(1964) reported that their Thai speakers used lead, short lag and long lag respectively for voiced, unaspirated voiceless and aspirated voiceless stops.

Across languages, the same consonant can demonstrate notable VOT differences in its phonetic expression. For instance, the mean VOT values for voiceless consonants are much lower in Thai or Korean than in English. However, common characteristics can be observed in the data across language. In most languages, the VOT values for voiced and voiceless stops are produced in discrete duration ranges that correspond to the voicing categories. That is, the categories of VOT values are separated by a range of times in which no production occurs. This categorization indicates active control for the voiced and voiceless cognates in the timing relationship between the articulatory release and the onset of vocal fold vibration. The articulatory gestures for short voicing lag stops are easier than for the other two types of VOT. Voicing lead requires more muscles gestures than those needed for short voicing lag stops. The production of initial stop consonants with voicing lead necessitates some mechanisms external to the larynx to sustain an adequate transglottal pressure drop during the closure. In contrast, the stop consonants with long voicing lag require more carefully controlled timing between the oral stop and
laryngeal closure. For these consonants, the adduction of the vocal folds requires more complex muscle activity (Dixit, 1989; Tyler and Watterson, 1991; Vohr, Garcia-Coll and Oh, 1988; Westbury and Keating, 1986). In producing the voicing contrast, a speaker needs to coordinate the timing of velopharyngeal closure, supraglottal articulators closure, vocal fold oscillation and supraglottal articulator release.

Another cross language commonality is that VOT varies regularly with the place of articulation in all languages. The lag time values tend to increase as the position of occlusion moves posteriorly within the oral cavity: the mean VOT values for [p] are shorter than for [t] which, in turn, are shorter than for [k] (figure 6) (Baum and Ryan, 1993; Klatt, 1975; Miller, Green and Reeves, 1986; Volaitis and Miller, 1992). However, the range of VOT values overlap extensively across places of articulation for both voiced and voiceless consonants.

Factors that can alter VOT production

Changes occur in VOT as a consequence of physiological differences (such as age, lung volume), pathological status (hearing impairment, depression) and different linguistic tasks (speech task, speech rate, phoneme environment). One such factor that has been investigated is the age of the speaker. Variations in VOT have been reported across the life span. Eguchi and Hirsh (1969) reported high VOT variability
Table 1. Normative values of VOT in English literature in milliseconds (means and standard deviations or range, if available). Some of them are approximated from figures. Lisker and Abramson (1964) gave two ranges of values for voiced stops corresponding respectively to lead and short lag (see text). C: Consonant; V: Vowel

<table>
<thead>
<tr>
<th>Authors</th>
<th>Number of subjects</th>
<th>p</th>
<th>t</th>
<th>k</th>
<th>b</th>
<th>d</th>
<th>g</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lisker and Abramson, 1964</td>
<td>4</td>
<td>58 (20:120)</td>
<td>70 (30:105)</td>
<td>80 (50:135)</td>
<td>101 (−130:−20); 102 (−155:−40); 88 (−150:−60);</td>
<td>1 (0.5)</td>
<td>5 (0.25)</td>
<td>21 (0.35)</td>
</tr>
<tr>
<td>Klatt, 1975</td>
<td>3</td>
<td>47</td>
<td>65</td>
<td>70</td>
<td>11</td>
<td>17</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Sweeting and Baken, 1982</td>
<td>30</td>
<td>76.6 (±4.6)</td>
<td>13.9 (±1.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardcastle et al., 1985</td>
<td>4</td>
<td>65.3</td>
<td>80</td>
<td>82.8</td>
<td>16.9</td>
<td>15.7</td>
<td>21.3</td>
<td>Single word</td>
</tr>
<tr>
<td>Morris and Brown, 1987</td>
<td>25</td>
<td>70</td>
<td>65</td>
<td>10</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caruso and Burton, 1987</td>
<td>8</td>
<td>62.5 (25.4)</td>
<td>71.9 (19.4)</td>
<td>74.8 (74.8)</td>
<td>19.7 (8.5)</td>
<td>21.4 (4.9)</td>
<td>35.2 (8.4)</td>
<td>Say CVC again</td>
</tr>
<tr>
<td>Lee et al., 1988</td>
<td>10</td>
<td>72.5 (26.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forrest et al., 1989</td>
<td>8</td>
<td>46 (10)</td>
<td></td>
<td></td>
<td>12 (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baum and Ryan, 1993</td>
<td>10</td>
<td>85</td>
<td>95</td>
<td>110</td>
<td>10</td>
<td>15</td>
<td>30</td>
<td>Twenty-one say CVC</td>
</tr>
<tr>
<td>Brown et al., 1993</td>
<td>12</td>
<td>57</td>
<td>65.2</td>
<td>6.4</td>
<td></td>
<td>16.8</td>
<td></td>
<td>Speak Cat again</td>
</tr>
<tr>
<td>Petrosino et al., 1993</td>
<td>10</td>
<td></td>
<td>89 (20)</td>
<td></td>
<td></td>
<td></td>
<td>14 (24)</td>
<td>Say Cat again</td>
</tr>
<tr>
<td>Davis, 1995</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
in preadolescent children speakers of American English. These authors speculated that the extremely precise temporal coordination involved in VOT requires a fully mature nervous system. In contrast, Ryalls and Larouche (1992) found comparable VOT values between children and adult speakers of Québec French. However, the differing results from these two studies may reflect a difference in the VOT categories used for phoneme contrast within the language.

Three steps in the developmental sequence for the acquisition of the voicing contrast have been described. The stops first appear around six months of age, with VOT values randomly distributed from voicing lead to long voicing lag. At one year of age, the stop productions separate into the appropriate phonemic categories (Tyler and Watterson, 1991). Then, the children modify their productions toward adult VOT ranges for voiced (short lag in English) and voiceless stop consonants (long lag in English). Most English-speaking children develop the voicing contrast by approximatively 30 months of age (Kewley-Port and Preston, 1974; Macken and Barton, 1980a). Surprisingly, a comparable study of Spanish-learning children showed that they acquire the Spanish voicing contrast after the age of 4 years (Macken and Barton, 1980b). The developmental pattern among young children indicates that short voicing lag stops are easier to produce successfully than the two other categories. No consistent gender based difference in VOT has been found at any age level (Sweeting and Baken, 1982; Zlatin, 1974).

As stated above for children, greater VOT variability has been reported for older adults. The mean VOT values for labial (Sweeting and Baken, 1982) and velar plosives (Petrosino et al., 1993) are similar in young and aged male subjects, therefore the older subjects produced VOTs that remained within the appropriate VOT categories. However, for the voiceless plosives, the older subjects produced shorter mean VOT’s for [p] among the male subjects (Sweeting and Baken, 1982) and for [p] and [t] among the female subjects (Morris and Brown, 1987). Thus, the older subjects exhibited reduced VOT differences between the voiced and voiceless stops. Finally, the older male and female subjects exhibited greater VOT variability than did the younger subjects (Morris and Brown, 1994; Sweeting and Bacon, 1982).
In normal speech production, VOT values vary with the speaking rate. VOT decreases as the rate increases, indicating that the overall rate changes are implemented at the segmental level (Baum and Ryan, 1993; Diehl, Souther and Convis, 1980; Jäncke, 1994; Miller, 1981; Miller et al., 1986). A slower speaking rate results not only in longer VOT but also in a wider range of VOT values (Miller and Volaitis, 1989; Volaitis and Miller, 1992). This effect is particularly prominent in English for voiceless stop consonants while values for voiced stops remain relatively stable. Thus, the difference between VOTs of voiced and voiceless consonants is reduced when the speaking rate increases. The changes in VOT values are also more marked on velar consonants than for apical or labial consonants. Thus, consonants with longer VOT values, velar and voiceless stops, permit greater decreases when speaking rate increases (Baum and Ryan, 1993). However, normal subjects maintain bimodal VOT distributions for all places of articulation, clearly separating the voiced and voiceless consonants (Baum and Ryan, 1993). Moreover, the modifications of VOTs induced by changes in speaking rate are more marked for aspirated than for unaspirated stops (Miller et al., 1986; Pind, 1995, 1996).

Measurements have been made of VOT in English plosives followed by vowels in word initial, medial, and final positions using both stressed and unstressed syllables. Few studies report the VOT values in consonant clusters. Klatt (1975) showed that VOTs were longer before sonorant consonants than before vowels. They were also longer before the high vowels [I] and [u] than before mid- or low vowels [ae] and [a]. In some way, the voicing seems to be more difficult to initiate when the laryngeal frequency is high. The high vowels are produced with greater longitudinal vocal fold tension due to a greater contraction in the cricothyroid muscle (Löfqvist, Baer, McGarr and Story, 1989). The increased vocal fold tension has been associated with longer VOT values (Löfqvist, 1992).

The physiologic interaction of the larynx and the respiratory system can affect VOT values. Hence, the VOT of [p] is longer at high lung volume and shorter at low lung volume in normal subjects (Hoit et al., 1993). For example, at high lung volumes, the diaphragm usually flattens and pulls the trachea and larynx caudally, exerting a force that tends to abduct the vocal folds (Zenker, 1964). This link between VOT and lung volume must be kept in mind when studying VOT in subjects exhibiting aphasia, apraxia of speech or dysarthria. The abnormally long or short VOT values produced by subjects with neurogenic speech problems are interpreted to reflect impaired coordination between the larynx and the upper airways. In fact, some disorders such as PD are also associated with the use of abnormal lung volume ranges during speech production (Murdoch, Chenery, Bowler and Ingram, 1989; Solomon and Hixon, 1993).

Changes in the VOTs produced by a group of mountain climbers who were recorded while they were ascending Mt Everest have been interpreted similarly (Lieberman, Protopapas and Kanki, 1995). These authors found that the mean interval that differentiated the VOTs of their voiced and voiceless stop consonants decreased at increasing altitudes from Base Camp to Camp Three. They concluded that smaller lung volumes were available to the climbers in the thinner air at higher altitudes. Thus, their findings agree with those of Hoit et al. (1993).

As opposed to the high altitude differences in healthy subjects, obstructive pulmonary disease does not appear to affect VOT. No differences were found between healthy and asthmatic subjects for the VOT of word initial [t] (Lee, Chamberlain, Loudon and Stemple, 1988). Additionally, Lee et al. reported that in
their healthy subjects, VOT was significantly reduced in the initial [t] of ‘21’ when counting loudly in comparison to counting at normal intensity.

Finally, VOT patterns have been studied in people exhibiting depression and hearing impairment. Depressed patients have significantly shorter VOT values compared to normals (Flint, Black, Campbell-Taylor, Gailey and Levinton, 1992). This finding is of interest when considering neurological disorders in which depression is encountered, for example Parkinson’s disease. Two studies report that speakers born deaf or deafened before learning English present a weaker distinction between voiced and voiceless consonants (Lane et al., 1994; Monsen, 1976). The subjects in these studies did not use longer VOT durations for the voiceless consonants. For French speaking children, Ryalls and Larouche (1992) found no difference between hearing-impaired children and age-matched subjects for the six stop consonants. As before noted, this cross language difference in results may reflect a difference in the categories of VOT used for making phoneme contrasts within the languages.

**Aphasia**

VOT has been measured in aphasic patients to help differentiate their phonemic and phonetic error patterns (Baum and Ryan, 1993; Blumstein, Baker and Goodglass, 1977; Itoh et al., 1982; Ryalls et al., 1995). One way that the VOT production of the aphasic patients matched normal speakers was that both fluent and nonfluent aphasics maintained the normal pattern of longer VOT values for velars in comparison to labial and alveolar stops (Baum and Ryan, 1993). However, several studies have found VOT differences between phonemic and phonetic errors. A phonemic error, phoneme substitution, is reflected by a VOT value lying within the normal range of the corresponding cognate. In contrast, a phonetic error phoneme-distortion, is reflected by a VOT value lying between the normal ranges of VOT for the voiced and voiceless categories. Blumstein (1980) provided the following example of the two error types. In English the normal VOT range of the voiced bilabial [b] is from $-105$ to $+15$ msec and the VOT range of its cognate [p] is from $+35$ to $+105$ msec. For a [b], a VOT greater than $+35$ msec falls in the VOT range of [p], and would be a phonemic error, whereas a value between $+15$ and $+35$ msec would be a phonetic error.

Using these criteria, Blumstein (1980) compared the VOT productions from subjects exhibiting three types of aphasia. She reported that the Broca’s aphasics made the greatest number of errors, they made errors during 40% of their VOT productions. For these speakers, 65% of the errors were phonetic. The conduction and the Wernicke’s aphasics made errors during 29% and 8% of their VOT productions, respectively. Even among the fluent aphasics who exhibited lower error rates, at least 50% of the errors were phonetic in nature. Several studies indicate that Broca’s aphasics present a continuum of VOT values with overlap between the short and long lag bilabial phonemes (Baum and Ryan, 1993; Blumstein et al., 1977; Blumstein, 1980; Gandour and Dardarananda, 1984; Gandour, Ponglropisit, Khunadorn, Dechongkit, Boongird and Boonklam, 1992). Other studies show that despite producing VOTs that occurred between the ranges for [b] and [p], Wernicke’s aphasics maintained a bimodal VOT distribution. The preserved separation of the VOTs for the cognate pairs indicates that the Wernicke’s aphasics better maintained this phonetic distinction (Baum and Ryan, 1993; Blumstein et al., 1977; Itoh et al., 1982; Shewan, Leeper and Booth, 1984; Tuller, 1984). Hence, the results for fluent
aphasics are interpreted as deficits of phonological planning rather than an articulatory impairment. On the contrary, the phonetic error patterns of the nonfluent aphasics correspond to deficits in temporal motor programming. Thus, Blumstein (1980) concluded that all of her aphasics exhibited deficits in control of the timing between their laryngeal and supralaryngeal articulatory adjustments. However, the fluent aphasic subjects (both Wernicke’s and conduction aphasics) exhibited timing control problems that were mild enough so that the voiced and voiceless productions were still distributed into relatively discrete VOT ranges.

In contrast, a comparable study using French aphasic patients failed to find more phonetic errors among Broca’s aphasics than among fluent aphasics (Ryalls et al., 1995). In addition, Buckingham and Yule (1987) and Gandour et al., (1992) suggested that some of the phonemic substitutions produced by fluent aphasics could reflect subtle motor disturbances rather than impairments in phonological planning.

**Apraxia of speech**

Few researchers have studied VOT in patients with apraxia of speech (Freeman, Sands and Harris, 1978; Hardcastle, Barry and Clark, 1985; Itoh et al., 1982). Apraxia of speech has been defined as a disorder of motor programming for speech with substitutions that tend to be variable and unpredictable (Darley, Brown and Swenson, 1975). Instrumental analyses have shown a high proportion of phonetic distortions in the speech of these patients, particularly in areas such as timing and coordination (Itoh and Sasunuma, 1984; Kent and Rosenbek, 1983) which might lead to some troubles in VOT production. Previous studies using fiber-optic and computer-controlled X-ray microbeam systems clearly indicate a defective temporal organization of articulatory movements in one apraxic patient (Itoh, Sasunuma, Hirose, Yoshioka and Ushijima, 1980; Itoh, Sasunuma and Ushijima, 1979).

As was reported for nonfluent aphasics, some authors report that apraxic subjects exhibit a considerable overlap in VOT distribution between long lag and short lag cognates (Freeman et al., 1978; Itoh et al., 1982). In contrast, Hardcastle et al. (1985) found a clear separation of long lag and short lag stops at the alveolar and velar places of articulation but a slight overlap at the bilabial place in one apraxic patient. For this patient, both the bilabial and alveolar long lag stops showed considerable variability in VOT values. The authors interpreted this general pattern as a problem in sensorimotor programming of speech with errors characterized primarily as selection, sequencing and temporal integration of target articulations. Spatial and temporal variability in achieving the articulation goals for this apraxic speaker seemed more marked than the low level of spatial and temporal variability frequently observed among dysarthric patients.

**Dysarthria**

Dysarthria is a term referring to a group of speech disorders resulting from damage to neural mechanisms that regulate speech movements. The classification of the dysarthrias according to the physiological approach of Darley et al. (1975) associates spastic dysarthria with lesions of the upper motor neurons, flaccid dysarthria with lesions of the lower motor neurons, ataxic dysarthria with cerebellar disease, hypokinetic dysarthria with extrapyramidal disorders such as Parkinson’s disease, and
hyperkinetic dysarthria with extrapyramidal disorders such as choreoathetosis or dystonia. When several of these systems are affected, the dysarthria is called mixed.

Several studies have included VOT measurements of the speech produced by subjects exhibiting different types of dysarthria. The speech samples used have included both oral reading and spontaneous speech tasks. Since Brown and Docherty (1995) found no consistent effect of sampling task (reading or spontaneous speech) on VOT of dysarthric patients, the results from these studies should be comparable.

Morris (1989) reported the VOT values for the three voiceless stops [p], [t] and [k] produced during repeated syllable tasks. He analysed audio tapes from 20 patients representing four types of dysarthria: five spastics, five flaccids, five ataxics and five hypokinetics; however, he did not include a control group. The results from this study are summarized in figure 7. He found that the mean VOT values increased as the position of occlusion moves posteriorly, with much overlap of values among the three consonants. When comparing the mean VOT values across dysarthria types, he found that the VOTs produced by the spastic dysarthrics for [t] were significantly shorter than those produced by the flaccid and ataxic dysarthrics. Similarly, Hardcastle et al. (1985) reported that spastic patients produce shorter VOTs than normally speaking control subjects.

Increased variability in the VOTs produced by dysarthric subjects has been reported in several studies. In his comparison of subjects exhibiting different dysarthrias, Morris (1989) reported that the patients with flaccid and ataxic dysarthria exhibited significantly greater VOT variability than those with spastic and hypokinetic dysarthria. In the flaccid dysarthria group, the interspeaker VOT variability was great whereas intraspeaker VOT variations were similar to those produced by spastic and hypokinetic dysarthrics. In contrast, the speakers with ataxic dysarthria exhibited not only interspeaker but also intraspeaker VOT variability. Other studies have confirmed the finding of greater VOT variability among ataxic patients (Ackermann and Hertrich, 1993; Keller, Vigneux and Laframboise, 1991). Similar VOT variability was also found among patients with Huntington’s disease, a hyperkinetic neurogenic disorder patients (Hertrich and Ackermann, 1994) and stuttering adults, even during non stuttering periods (Jäncke, 1994). In their study of eight

![Figure 7](image-url)  
Figure 7. Mean and standard deviations of each voiceless stops according to the type of dysarthria type (adapted from Morris, 1989).
patients with amyotrophic lateral sclerosis (ALS) and eight age-matched controls, Caruso and Burton (1987) reported no significant differences among the mean VOTs of the six stop consonants. However, the variability of the VOT was greater in the ALS group for all of these consonants except [p].

Results concerning the VOTs produced by subjects who have PD vary among the studies. Forrest, Weismer and Turner (1989) compared PD patients with age-matched normal subjects. They measured the VOT for the bilabial stops (five [b] and one [p]) in sentences such as 'Build a big building' or 'Buy Bobby a poppy'. The PD speakers produced longer VOTs than control speakers for all five [b] productions, but only the utterance initial [b] productions were significantly longer. The authors interpreted this as representing movement initiation problems at the laryngeal level. No significant difference occurred for the [p] VOTs in 'poppy', even though these VOTs tended to be longer than those produced by the normal speakers. This lack of significant difference for the VOTs of long lag stops between normal and PD speakers was confirmed in two studies of isolated and repeated CV syllables (Cohen, Laframboise, Labelle and Bouchard, 1993; Connor et al., 1989). In a study of French speaking subjects the VOTs of [p], [t], and [k] produced during a repeated syllable task by 18 PD patients were found to be significantly longer than those produced by 12 normally speaking control subjects (Özsancak, Auzou, Jan, Léonardon, Gaillard and Hannequin, 1997). In contrast to the reports that subjects with PD use longer VOTs, Flint et al., (1992) found significantly decreased VOTs in patients with PD versus normally speaking controls for initial [p], [t], and [k] consonants. This finding is unique among the VOT studies of subjects with PD.

Since the French subjects with PD exhibited significantly longer VOTs for [p], [t] and [k] than the normally speaking control subjects and the VOT differences for the English speakers with PD were significantly longer only for [b], the cross language results from PD patients appear to be contradictory. However, the contradiction is present only if the plosives are considered in terms of the voiced–voiceless contrast. The results can be more easily interpreted in term of the short voicing lag-long voicing lag contrast (table 2). In French, the voiceless stops use the short voicing lag and the voiced stops use the voicing lead. Thus, the French voiceless stops are produced with similar VOTs as the English voiced stops. We can then summarize the data from the previously reported studies as follows: in subjects with PD, the short lag durations are longer than normal and the long lag durations are normal or shorter. The reduction of the timing interval between these two VOT categories may lead to an overlap between the two stop consonant cognates and the perception of phonetic errors, as mentioned by Lieberman, Kako, Friedman, Tajchman, Feldman and Jiminez (1992). As Tyler and Watterson (1991) speculated, it appears that the excess muscular tension exhibited by PD speakers inhibits appropriately

<table>
<thead>
<tr>
<th>Authors</th>
<th>Short lag</th>
<th>Long lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forrest et al., 1989</td>
<td>increase</td>
<td>N</td>
</tr>
<tr>
<td>Connor et al., 1989</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Flint et al., 1992</td>
<td>decrease</td>
<td>N</td>
</tr>
<tr>
<td>Lieberman et al., 1992</td>
<td>overlapping</td>
<td>N</td>
</tr>
<tr>
<td>Cohen et al., 1993</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Özsancak et al., 1997</td>
<td>increase</td>
<td></td>
</tr>
</tbody>
</table>
timed approximation and initiation of vibration, leading to longer short lag durations. To some extent the VOT modifications in PD might also result from a reduction in the coordination among the different structures in the speech production mechanism.

**Conclusion**

Instrumental analysis provides quantitative and objective data on a wide range of speech parameters and greatly enhances the scope of auditory-based perceptual judgments of speech. Our review suggests that VOT is a reliable temporal parameter when appropriate methodological precautions are followed, such as using time-synchronized wideband spectrographic and oscillographic displays.

VOT reflects the temporal control between the larynx and the lips, tongue and jaw. It can be used to define the phonetic voiced–voiceless contrast and has been a functional tool for exploring the acoustic aspects of some speech disorders. When the subjects have had neurologically based communication disorders, researchers have varied their use of VOT depending on whether the subjects exhibited aphasia or dysarthria. When the subjects exhibited aphasia, the authors usually reported the distribution of VOT values along a continuum rather than the mean values within a voicing category. When so used, these data indicated the clinical distinction between phonetic errors, corresponding to a deficit in temporal motor planning, and phonemic errors reflecting a deficit of phonological planning. However, some authors suggest that part of the phonemic substitutions produced by fluent aphasics could also reflect subtle motor disturbances rather than impairments in phonological planning.

When the subjects exhibited dysarthria, the goal was to evaluate articulatory impairments and VOT productions have mainly been interpreted by using mean values or standard deviations. When VOT abnormalities are present, they mainly correspond to a shortening of long lag durations or a lengthening of short lag durations. These alterations are usually interpreted as a loss of timing control, but other variables such as the size of the glottal opening, the transglottal pressure and the vocal fold tension can also affect the VOT values. For example, the VOT values may be influenced by the respiratory impairment typically found among people with PD. In addition, the variability of VOT productions has been shown to be an important factor to consider in speakers with several types of dysarthria. Even when normal mean values occur, increased variability has been found in ataxic and hyperkinetic dysarthric subjects.

Our analysis of the literature leads us to use descriptive statistical measures of variability to complement the mean VOT values, such as standard deviations or the coefficient of variation. In addition, using graphs to represent the distributions of the VOT differences between the voiceless and voiced stop cognates often clarifies the data patterns. Previous studies also show the importance of measuring VOT at different speaking rates. The combined use of these different methodological tools that may allow the distinction of specific diagnostic patterns among the different dysarthria types.

VOT could also be used as an indirect mean to study vocal tract motor control. As speech reflects axial control, the measure of VOT could be used to compare axial with limb motor control. Such comparisons would be of great interest, for instance in PD, because these two types of motor control are not influenced in the same
manner by dopaminergic drugs. VOT could be then used as an objective parameter to evaluate the effect of new drugs on parkinsonian dysarthria in comparison to limb tremor.

Cross language studies using groups of subjects matched for diagnosis and severity of communication disorder could help to clarify the VOT variations observed between languages. Studies of this type will help researchers to better characterize the common acoustic abnormalities underlying disordered speech production. For this purpose, we believe it is relevant to consider the consonants in terms of ‘lead’, ‘short lag’, and ‘long lag’ rather than voiced and voiceless as this system better describes the VOT categories used across different languages.

Finally, current physiological theories about VOT production indicate interactions between it and the physiologic adjustments that occur at the different levels of the speech production mechanism. Further studies are needed to establish definitive normative data for these interactions. In addition, further studies should provide descriptions of the relationships between the clinical differences in speech production encountered as a result of motor disorders at the respiratory, phonatory and articulatory levels and their acoustic consequences.

References


