1197

Interdisciplinary Approaches to Language Processing

Editors

Denis Burnham, Sudaporn Luksaneeyanawin,
Chris Davis, and Mathieu Lafourcade

The Perception of Voicing Distinctions

Arthur S. Abramson

Abstract

The term *voicing* is used as a label for both a phonetic property and a phonological feature, sometimes leading to confusion. The most salient phonetic aspect of voicing is audible glottal pulsing. Voicing in a part of an utterance interests us here only if its presence is in opposition phonologically to its absence. Although glottal pulsing is the dominant excitation source in speech, there are also such noise sources as turbulence and transients. In different contexts and across languages, phonological voicing distinctions entail various combinations of these sources as well as such concomitant traits as differences in fundamental frequency upon consonant-release, preconsonantal vowel duration, and intervocalic closure duration. Much perceptual research has been done with synthetic speech and manipulated natural speech. In my earliest research with Leigh Lisker on the acoustics of voicing distinctions, it was convenient in working with a variety of languages to focus on initial position, so the concept of voice onset time (VOT) came to the fore. This is the time of the onset of voicing relative to the release of the initial consonant. Actually, the broader concept of voice timing is relevant to initial, word-medial, and utterance-final positions. Although voice timing by itself is a powerful mechanism for perceptual differentiation of voicing states, research has shown that the concomitant traits mentioned above can also play a role in perception. Voice timing is broadly applicable in languages of the world; yet there are some languages in which non-temporal characteristics intersect with that dimension and thus must be handled and processed separately.

1. Background

The word *voicing* is meant to be a technical term in linguistics and speech research, yet it is beset with a certain amount of confusion. The phonologist working within a particular theoretical framework may use the term as the label for an abstract phonological feature that is said to play a distinctive role in a grammar. The field phonetician may take it to mean the presence in a portion of a speech signal of audible glottal pulsing. The laboratory phonetician may even wish to label as voiced a span of speech with glottal pulsing that is instrumentally detectable but too weak to be audible. The confusion I have in mind has come about mainly on the part of phonologists whose apparent conviction that a phoneme is manifested in an utterance only in a well-defined segment leads them to reject the relevance of any feature not present in *that* segment. I will come back to this point in a later section. Before

reviewing findings on the perception of voicing distinctions, it may be helpful to have a brief overview of excitation sources in speech.

2. Excitation Sources in Speech

The acoustic consequences of speech gestures are modulated upon some kind of carrier that makes the phonetic information audible. In normal speech the carrier is a series of regularly spaced pulses from the opening and closing borders of the glottis, the vocal folds. This is laryngeal phonation, i.e., voice. The second source, an aperiodic one, is noise, which occurs as turbulence or a transient. There are two kinds of turbulence in speech. Air under sufficient subglottal pressure pushed through a suitable glottal opening will stir up eddies of noisy turbulence at the glottis. While it lasts, this glottal turbulence, generally heard as aspiration, is the carrier of the speech signal. Another kind of turbulence is local frication, which occurs when the air stream is forced through a narrow constriction somewhere in the tract above the glottis, as in fricative consonants. Finally, there is the possibility of a transient, a shock-excitation of the vocal tract caused by the sudden release of air under pressure from behind a closure, the so-called "burst" of a stop-consonant release. To these it is useful to add, for phonetic purposes, something that I think is unconventional in physical acoustics, namely, a null source. In speech there are frequent brief silent gaps as parts of articulatory gestures. Although there is no sound source in such a gap, that gap does carry phonetic information.

We know of no spoken language in which voice is not the normal carrier. No doubt language has evolved this way everywhere because of the efficiency of the voice as a carrier. The varying resonances of the rapidly changing configurations of the supraglottal tract are best excited by a periodic wave rich in harmonics such as glottal pulses or those from an electrolarynx. Languages, however, have exploited the possibility of switching between excitation sources for either cultural or linguistic reasons. In whisper—true whisper, that is— only noise sources are used. In murmur and breathy voice there is a mixture of turbulence and voice. These largely socio-cultural functions are different from the switching between sources for phonological distinctions.

Commonly, for a subclass of the consonants of a language the members occur in voiced and voiceless pairs.[1] That is, for each of those pairs the phonology dictates that the normal voice carrier is turned off in the "voiceless" member and replaced by either a noise source or silence for a crucial part of the articulation. The silence is one of those gaps just mentioned, as in the stop closures of such English words as *spill, still, skill* or the closure of the initial stop in a Thai word like /ta:/ 'eye' with, as it is traditionally described, an initial voiceless unaspirated stop.

Now and then one hears of a language with a voicing distinction in its vowel system,[2] but this is rare, perhaps because most syllable nuclei are vowels and the carrier is best radiated and transmitted during the production of vowels, especially if it is voiced. On the other hand, note that "aspiration" can well be understood as excitation of a portion of a vocalic span by glottal turbulence. What we transcribe as [h] and treat phonologically as a consonant phoneme, is nothing but aspiration as the sound source for the beginning or ending part of a vowel. This mechanism also covers the aspiration described for some of the stop consonants in many languages, whether pre-

aspirated as in Icelandic (Pind, 1995) or post-aspirated as in Thai (Lisker & Abramson, 1964).

Various combinations of the sources take place in speech. In a Thai word like /du:/ 'to look at' or a French word like /du/ *doux* 'sweet' voicing starts during the stop closure well before the release and continues through the vowel; the burst of the release occurs in the train of glottal pulses. In voiced fricatives, as in the medial /z/ of English *easy*, much of the glottis is in vibration for voice even while a portion of its length is kept open enough to furnish an air stream for a local constriction. A similar laryngeal adjustment is used for murmur or breathy voice. It should be understood, by the way, that whether to view a sequence of sources as a combination within one segment is more a phonological judgment than a phonetic one.

It has long been observed (e.g., House & Fairbanks, 1953) that various acoustic properties are often found in conjunction with voicing distinctions. The three outstanding ones have to do with vowel duration, closure duration, and fundamental-frequency (F_0) perturbations. There is a general tendency for vowels to be longer before voiced consonants. This is especially true in English where, it has been argued (e.g., Kluender, Diehl, & Wright, 1988), it may have been enhanced diachronically for auditory effects. There is some electromyographic evidence of motor control of vowel-lengthening before voiced consonants (Raphael, 1974). In intervocalic stop consonants closure durations are greater for voiceless stops. Upon release of a prevocalic voiceless stop, the F_0 of the onset of the voice source is likely to be higher than after a voiced stop (Lehiste & Peterson, 1951; Kohler 1982; Umeda, 1981; Ohde, 1984). Many attempts have been made to explain this phenomenon. To me the most convincing explanation is as a consequence of the role of the cricothyroid muscle in helping to suppress phonation (Löfqvist, Baer, McGarr, & Story, 1989).

3. Acoustic Cues to Voicing Distinctions

As suggested in Section 1, apparently in the strong belief that a phoneme manifests itself in speech in a very narrowly defined window, the "segment," some linguists rejected voicing as a phonologically relevant feature because it did not reliably appear in such segments as they took to be some of the contextual variants (allophones) of putative "voiced" consonants. That is, given the premise of the segment, they concluded that some other feature must distinguish members of the phonemic category from its allegedly voiceless counterparts. Thus it came about that the opposition "fortis–lenis" or "tense–lax" came into being for English, German, and some other languages. By this reasoning, voicing, when it did occur in the "lax" category, came about as a secondary or concomitant effect of the lower level of articulatory effort (see, e.g., Jakobson & Halle, 1962). The perceptual assumption was apparently that the bundle of acoustic properties associated with each category, tense and lax, served as acoustic cues to the distinction.

I hasten to add here that being skeptical of the foregoing argument does not require the dismissal of the physiological possibility of using level of effort for phonological distinctions. For example, a language can use extra contraction of the thyroarytenoid muscle for systematic shifting of voice quality in the vowel following the release of members of a particular consonant class; the phonologist might then reasonably invoke a feature of tensity. In the case of the absence of voicing in certain "voiced"

segments, however, those phonologists leapt to a conclusion without good phonetic evidence.

Leigh Lisker and I have argued that, at least until recently, phonological theories have spurned temporal control as a crucial factor in phonemic distinctions (Abramson & Lisker, 1970; Lisker & Abramson, 1971). Stimulated by the implications of the work of forerunners (e.g., Liberman, Delattre, & Cooper, 1958; Fant, 1960) and by our own auditory impressions, we collected acoustic data on the homorganic stop categories of 11 languages (Lisker & Abramson, 1964). We chose these languages as representative of types with two, three, or four such categories that, however they have been described in the literature, all seemed susceptible of differentiation to some extent by laryngeal timing. Our hypothesis was in fact that the dimension of relative timing would definitely fail to handle only the "voiced aspirates" of our two Indo-Iranian languages, Hindi and Marathi. We were also a little doubtful of complete success in Korean.

In this first study we focused on word-initial position, the only one in which none of the languages failed to maintain the distinctions in question. We measured the interval between the onset of glottal pulsing and the acoustic sign of the release of the stop in isolated words and words embedded in sentences. The release was assigned the value of 0 msec; voicing onset before the release, i.e., in the closure, was assigned a negative value and called voicing lead; voicing after the release was assigned a positive value and called voicing lag. With our focus at the time on initial position, we called the dimension voice onset time (VOT). It is illustrated in Figure 1 with the three categories of Thai that are conventionally called voiced, voiceless unaspirated, and voiceless aspirated respectively. We now regret not having used a more encompassing label for the concept, like "voice timing," to cover its relevance in other contexts (Abramson, 1977).

By and large our hypothesis was supported. In all the two category languages, such as English and Spanish, the categories were well separated by VOT, although the ranges and boundaries differed. In the three-category languages, including Thai, it worked well too, except for Korean in which two of the categories were not well separated from each other in initial position, although all three were separated by VOT in intervocalic position. The voiced aspirates of Hindi and Marathi were not distinguished from the voiceless aspirates. Each of these conflicts is resolved by the intersection of another non-temporal laryngeal dimension with VOT, tense voice for the first and murmur for the second.

Using a parallel-resonance synthesizer, we prepared stimuli varying in small steps of VOT from -150 msec through 0 to +150 msec for labial, apical, and dorsal stops. The continuum was complex because it simulated schematically the spectral changes that occur as voice onset moves from lead to lag. Thus, although we were using an acoustic synthesizer, we had in mind the shifting phase relations between the larynx and supraglottal articulators. We have used these stimuli, as have others (e.g., Neary & Rochet, 1994), in a number of studies (e.g., Abramson & Lisker, 1965, 1970, 1973; Lisker & Abramson, 1970) and found great perceptual efficacy of VOT for the several languages tested.

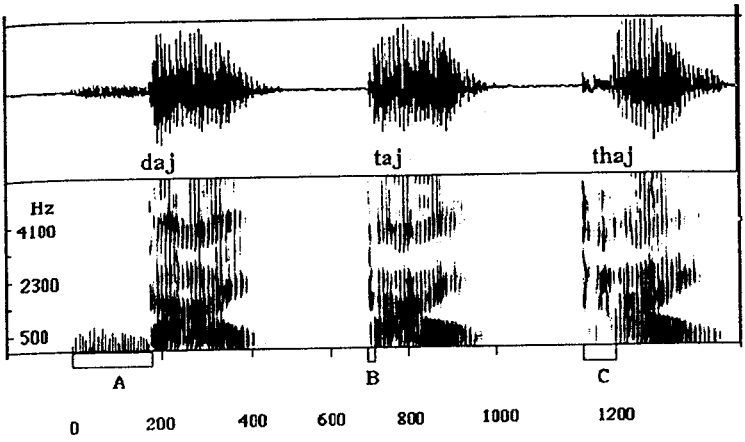

Figure 1. Thai illustration of voice onset time. A=voicing lead of -75 msec, B=short voicing lag of 5 msec, C=long voicing lag of 30 msec.

Other acoustic properties mentioned above have been found to serve as cues to the voicing distinction. In synthetic speech vowel duration (Denes, 1955; Raphael, 1981) is a sufficient differentiator when other factors are kept neutral for voicing distinctions in final consonants. F_0 shifts alone furnish only weak cues, but coupled with VOT variants, they have a significant effect on the boundaries between perceptual categories (Haggard, Ambler, & Callow, 1970; Whalen, Abramson, Lisker, & Mody, 1993).

English trochaic words (strong stress followed by weak stress) are particularly interesting because the intervocalic stops in them, at least in American English and some other dialects, are phonetically different from their counterparts in initial position. In trochees medial voiced stops are likely to have pulsing through their closures, while voiceless stops are unaspirated. It has long been known (Lisker, 1957; Port & Dalby, 1982) that longer closure durations will yield perceptual judgments of voicelessness and shorter ones, of voicing, as long as no pulsing is present in the closures of synthetic speech. We are now writing up an elaborate study of voicing in trochaic words with manipulated natural speech (Abramson, Koenig, & Lisker, in preparation). We have found three factors to be powerful cues: glottal pulsing, closure duration, and duration of the stressed vowel. We found differences in release-burst intensity, commonly assumed to be relevant, to have no value as a cue.

4. Conclusion

The timing of voicing relative to supraglottal articulatory gestures is a powerful differentiator of homorganic stop categories in initial and medial position. It appears to serve the purpose in the form of temporal offset in final position for many languages. Beyond VOT, the relative timing of vowels and stop closures is an important cue to voicing distinctions. Although historical linguistic evidence implies

that speakers of a language can become aware of these mechanisms (e.g., Maspero, 1911), as well as F_0 shifts, and enhance them, it is moot as to which of them may be under voluntary control.

---

[1] "Voicing" distinctions in sets of three or even four consonants of the same place of articulation will be handled below.

[2] Here I do not consider non-distinctive contextually determined "devoicing" of vowels as in common productions of the unstressed first syllable of English *potato*.

Acknowledgments

Much of the research described in this paper has been supported by two NIH grants to Haskins Laboratories, HD01994 and DC 02717. The University of Connecticut Research Foundation provided funding for the author's travel to Thailand to participate in the International Workshop on Human and Computer Processing of Language and Speech and to do related research there for three months.

References

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

Abramson, A.S., Koenig, L., & Lisker, L. (in preparation). Medial voicing distinctions in English trochaic words.

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Commins (ed.), *Proceedings of the Fifth International Congress on Acoustics, Ia* (A51). Liège.

Abramson, A.S., & Lisker, L. (1970). Laryngeal behavior, the speech signal and phonological simplicity. In *Actes du X^e Congrès International des Linguistes, 4* (pp. 123–129). Bucharest.

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569–573). Prague.

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1–8.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America, 27*, 761-764.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton. [2nd ed., 1970].

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47*, 613–617.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105–113.

Jakobson, R., & Halle, M. (1962). Tenseness and laxness. In *Roman Jakobson: Selected writings, I. Phonological studies* (pp. 550–555). The Hague: Mouton.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16*, 153-169.

Kohler, K.J. (1982). F_0 in the production of lenis and fortis plosives. *Phonetica, 39*, 199–218.

Lehiste, I., & Peterson, G.E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 419–425.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech, 1*, 153–167.

Lisker, L. (1957). Closure duration and the intervocalic voiced–voiceless distinction in English. *Language, 33*, 42–49.

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

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563–567). Prague.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language, 47*, 767-785.

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

Maspero, H. (1911). Contribution à l'étude du système phonétique des langues thai. *Bulletin de l'École Française d'Extrême-Orient, 19*, 152–169.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(1–19)

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75*, 224–230.

Pind, J. (1995). Constancy and normalization in the perception of voice offset time as a cue for preaspiration. *Acta Psychologica, 89*, 53–81.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics, 32*, 141–152.

Raphael, L. J. (1974). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics, 3*, 25–33.

Raphael, L. J. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica, 38*, 126–147.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *Journal of the Acoustical Society of America, 70*, 350–355.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F_0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America, 93*, 2152–2159.