CONTEXTUAL EFFECTS ON CONSONANT VOICING PROFILES: A CROSS-LINGUISTIC STUDY

Chilin Shih*, Bernd Möbius*†, and Bhuvana Narasimhan*
**Bell Labs – Lucent Technologies, USA †University of Stuttgart, Germany*

ABSTRACT

In this paper we present selected results from a study of the voicing profiles of consonants in five languages, viz. Mandarin Chinese, German, Hindi, Mexican Spanish, and Italian. We will focus here on the voicing properties of stop closures in these languages. The voicing profile is defined as the frame-by-frame voicing status of a speech sound in continuous speech. We propose statistical models that predict the probability of voicing from phone identity, neighbouring phones, and positional and prosodic factors.

1. INTRODUCTION

This paper investigates the voicing properties of stops by means of voicing profiles: the frame-by-frame voicing probability throughout the duration of stop closures. The voicing profiles reveal the dynamics of voicing status changes and thereby facilitate further investigations of contextual effects on voicing.

The motivation of this study originally comes from applications in speech technologies, such as speech synthesis, speech recognition, and automatic speech segmentation. Some applications in these domains are sensitive to the discrepancies between the assumed (often phonological) specification of a speech sound and its acoustic realization. The phonological specification of voicing, represented as a binary distinction of [+voice] or [-voice] over the domain of the entire speech sound [2] is a particularly troublesome feature when it comes to feature-to-data matching, when the state of vocal cord vibration is subject to the delicate balance of many factors [16, 10, 1, 15, 11, 9]. Such discrepancies are of course well-known, and have been a major motivation driving the research of voice onset time (VOT) [8, 7] in search of a better criterion than the distinctive feature to differentiate the stop series in a given language. The VOT research has been successfully applied cross-linguistically [4, 6, 14, 12, 3].

As more information becomes available in addition to VOT, such as the voicing trajectory along the entire stop closure duration, we expect a better understanding of the factors affecting voicing contrasts, and improvement in speech related applications. This is the starting point of the current paper. We will compare the voicing profiles of the stop closures in five languages, viz. Mandarin Chinese, German, Hindi, Mexican Spanish, and Italian, and discuss the effects of aspiration, preceding phone context and following phone context, and prosodic factors where available. The study is part of a larger project that investigates the voicing profiles of consonants in a cross-linguistic framework (see also [13]).

2. SPEECH DATA

The speech databases of the five languages were each produced by a single speaker. The Mandarin and Spanish sentences were chosen from online text corpora by means of a greedy algorithm, maximizing phone combinations and the interaction of phonemes, prosodic factors, and positional factors. The Italian corpus and the Hindi corpus contain target words embedded in sentence frames. The words occur in sentence initial, medial and final positions. The German database is a subset of the Kiel Corpus of Read Speech [5]. We analyzed stops and affricates from these databases. The number of sentences and the number of stops and affricates in each corpus is listed in Table 1. As a rule, the affricates exhibit patterns similar to those of the stops in each language.

The phone boundaries were labeled manually. One labeling criterion particularly relevant to this study is that the boundary between a vowel and the following stop was placed where the vowel formants, especially F_2, disappear, which is a good indicator that the stop closure is being formed.

Voicing information was obtained automatically using the ESPS/Waves speech analysis software of Entropic Inc. The program reports a binary voicing decision for each analysis frame with "1" representing voiced and "0" representing unvoiced.

3. ADDITIVE MODELS

We obtained 11 samples of voicing information from each stop closure duration at 10 equidistant time intervals, including the initial and final frames, and trained additive models for each position separately. We used the observed voicing information by position as the dependent variable and, minimally, the following factors as independent variables: phone identity; phone class of the preceding sound; phone class of the following sound; closure duration.

For Mandarin, German, and Spanish, where the databases consist of natural sentences, we also trained models using additional factors, including prosodic factors such as stress and accent, and positional factors such as whether a sound is in the initial, medial, or final position of the syllable, the word, the phrase, and the utterance. In addition, tonal information is used for Mandarin, and syllable onset and coda labels are used for Spanish and German.

In all following figures, coefficients of the factor levels are plotted on the y-axis as a function of normalized time, expressed as percentage of the closure duration. Coefficients of the relevant factor levels are summed up for each position on the normalized time axis to obtain the predicted voicing of a given phone in a given context. The voiced/voiceless threshold was set at 0.5.

Table 2 works through one example by calculating the voicing probability of the /p/ closure at six positions, in the segmental context of /aps/ in Spanish. Factor "phone identity" has seven levels representing 6 stops and 1 affricate; the /p/ closure is coded as "p" in Spanish (see Figure 1). Factor "preceding phone" has six levels; in our example, the preceding phone is /a/, coded as "V" (see Figure 2). Analogously, factor "following phone" also has six levels,

	# sentences	# stops/affr.
Mandarin	424	5569
Hindi	566	1319
Spanish	634	7010
Italian	1203	3231
German	598	4303

Table 1. Number of sentences and phones analyzed.

Position	0%	20%	40%	60%	80%	100%
PhoneID (p)	0.67	0.54	0.38	0.23	0.13	0.11
Prec Ph (V)	0.26	0.23	0.21	0.12	0.03	0.01
Foll Ph (s)	0.05	0.10	0.12	0.09	0.06	-0.04
Sum	0.98	0.87	0.71	0.44	0.22	0.08
Pred. voicing	1	1	1	0	0	0

Table 2. Voicing prediction of /p/ closure in Spanish /aps/.

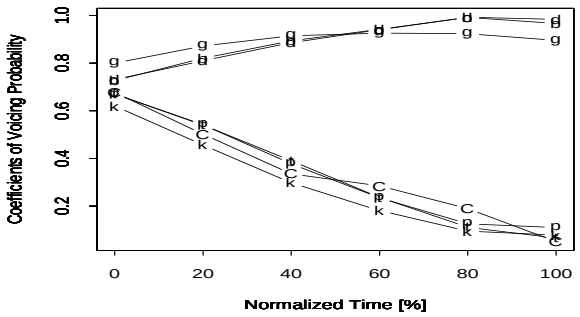

Figure 1. Spanish voicing profile: effects of phone identity. "C" represents the voiceless dental affricate.

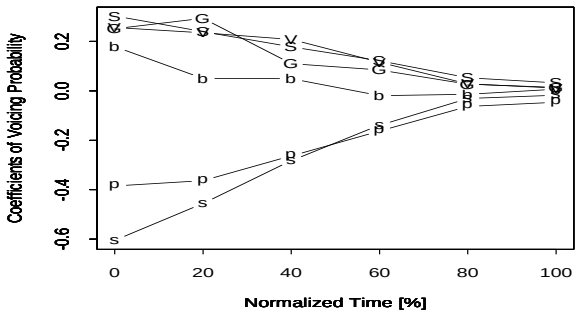

Figure 2. Effects of the preceding phone in Spanish. "V" = vowel, "G" = glide, "S" = sonorant, "s" = voiceless fricative, "p" = voiceless stop, "b" = voiced stop.

with "s" representing voiceless fricatives. For this combination of factors, the prediction is that, for instance, the probability is 87% for /p/ to be voiced at 20% into the closure duration, and only 22% at 80% into the closure.

In the languages under investigation, we see a clear separation of the stop series into a more voiced population and a less voiced population. In Spanish, Italian and Hindi, the separation tends to agree with the phonological specification of voicing. In Mandarin, the population is divided by the aspiration feature, even though phonologically speaking, both Mandarin stop series are voiceless. German voiced stops often do not have sustained voicing throughout the stop closure.

3.1. Spanish

Phone identity, preceding phone and, to a lesser extent, following phone are the three most important factors affecting stop closure voicing in Spanish. Figure 1 shows the coefficients for the factor phone identity. The voiced stops /b,d,g/ are clearly separated from the voiceless stops /p,t,k/ and the voiceless dental affricate /C/. The two classes of sounds are most clearly differentiated near the end of the closure duration, and most confusable in the beginning where the effect of the preceding phone is strong. Also note that the voiceless stops may have sustained voicing that extends far into the closure duration.

Figure 2 shows the effect of the preceding phone on the voicing of stops. The clearest effect is that if the preceding phone is a voiceless obstruent, there is a devoicing effect that extends up to 80% into the closure duration. The impact is strongest in the early part of the stops, as expected.

There is also a noticeable, albeit rather weak, effect of the following phone on the voicing of stops. We find very little impact from positional factors, and closure duration has no effect either, except in the latest position, where shorter duration increases voicing probability and durations of more than 100 ms decrease voicing probability.

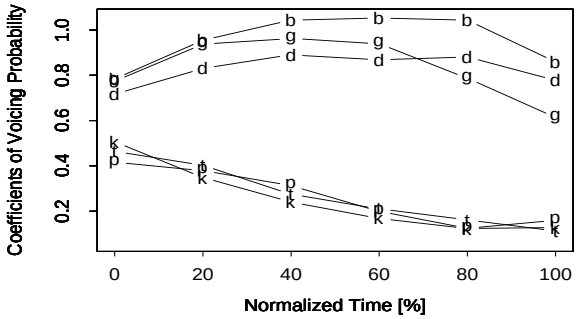

Figure 3. Italian voicing profile: effects of phone identity.

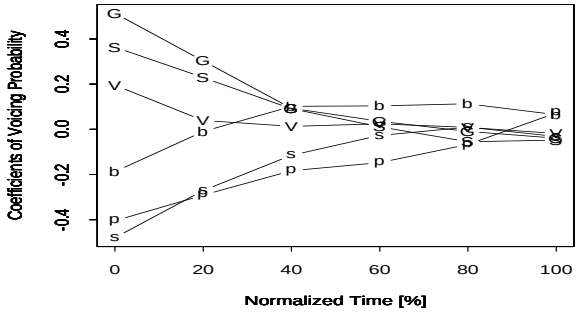

Figure 4. Effects of the preceding phone in Italian. Plotting symbols as in Figure 2.

3.2. Italian

The models for Italian single and geminate stops were trained separately. We noted that 100% of Italian geminates in our data agree with the phonological specification of voicing in the center 30–70% of the closure duration. The geminate plots are not shown.

The voicing patterns of Italian single stops in Figure 3 are quite similar to the patterns seen in Spanish. One difference is that in the latest position Italian voiced stops are more prone to devoicing.

Figures 4 and 5 show the coefficients of the preceding and following phones, respectively. Preceding voiceless phones have a devoicing effect on the early part of the closure, which is expected (Figure 4). However, preceding voiced stops, represented by the plotting symbol "b", pattern with preceding voiceless obstruents, suggesting that at least some samples of these canonically voiced context phones are in fact unvoiced. Also note that in Figure 5, following voiced stops have a voicing effect on the current phone from the 20% point on, and following voiceless stops ("p") have a devoicing effect from the 40% point on. Other sound categories have very little impact. The observed pattern suggests that there is a tendency for the voicing of an Italian stop to assimilate to the voicing specification of a following stop, a process that is reminiscent of the well-known consonant gemination process of Italian. Non-geminate stop-stop combinations are quite rare in Italian.

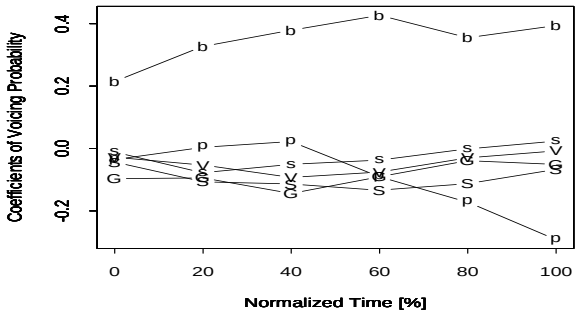

Figure 5. Effects of the following phone in Italian. Plotting symbols as in Figure 2.

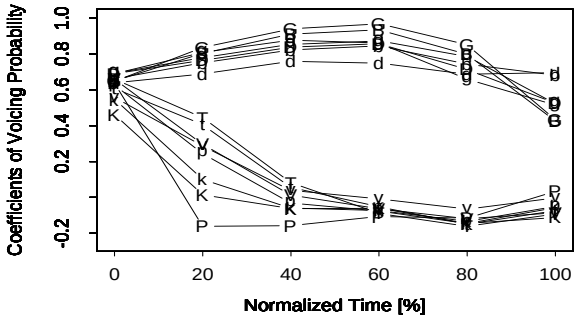

Figure 6. Hindi voicing profile: effects of phone identity. The phonetic voicing property of the stop closure corresponds well with the phonological specification: voiced (top) and voiceless stops.

3.3. Hindi

Hindi has a large paradigm of stops contrasting in voicing, aspiration and four places of articulation. Figure 6 shows the coefficients of the phone identity factor. First of all, the stop populations are divided by voicing. Phonologically voiced stops are indeed fully voiced, especially in the center, while phonologically voiceless stops turn voiceless by the 40% point. Thus, the phonetic voicing property of Hindi stop closures corresponds well with the phonological specification of the stops.

Aspiration has an interesting effect on voicing. Voiced aspirated stops /B, D, G, Q/ are more solidly voiced than their unaspirated counterparts /b, d, g, q/ (/Q, q/ represent voiced retroflex flaps) in the center, but the last frame right before the burst tends to be unvoiced. The same contrast in the last frame can be observed in, e.g., the waveform display in [14, p. 146] and the spectrograms in [6, p. 59].

Moreover, aspiration has a weak devoicing effect on voiceless stops. For most of the aspirated/unaspirated pairs, the aspirated one has a lower probability of voicing at a given position. This effect is most noticeable in the 20–40% region before all voiceless stops become completely unvoiced. Another way to describe this situation is that an aspirated voiceless stop becomes voiceless earlier than the corresponding unaspirated one, everything else being equal.

As for the impact of context phones on the voicing probability of Hindi stop closures, we observe only the expected assimilation effects, with preceding phones having a stronger influence than following phones.

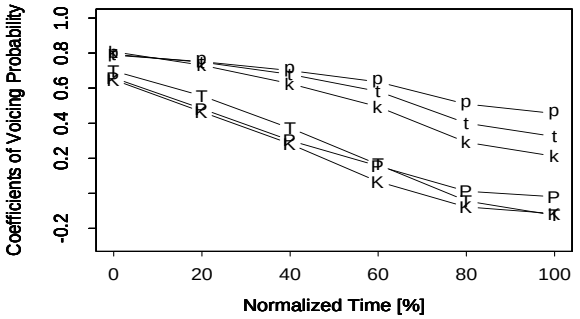

Figure 7. Mandarin Chinese voicing profile: effects of phone identity. Aspirated stops /P, T, K/ and unaspirated stops /p, t, k/ are separated along the voicing dimension.

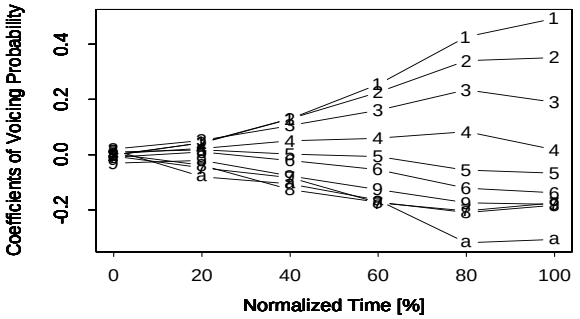

Figure 8. Effects of stop closure duration in Mandarin. “1”–”9” = closure durations of 10–90 ms, “a” = longer than 90 ms.

3.4. Mandarin

All Mandarin stops are phonologically voiceless. However, Figure 7 shows a complete separation of the aspirated and unaspirated population of stops along the voicing dimension. Uppercase plotting symbols indicate aspiration, as in Hindi.

In Mandarin Chinese, stops occur only in syllable initial position. Compounded with the fact that there are no consonant clusters and no syllable final voiceless consonants, it follows that Mandarin stops are always surrounded by voiced sounds, except in utterance initial position where the preceding sound is silence. These phonotactic constraints account for rather small variations in the effects of preceding phones, except for a strong devoicing effect of preceding silence. There is practically no influence from the following phone.

Stop closure duration has a clear effect in the second half of the closure (Figure 8): the shorter the duration, the higher the voicing probability. The strength of this effect increases toward the end of the closure.

3.5. German

German stop consonants are usually classified according to the voicing feature: /p, t, k/ are voiceless, /b, d, g/ are voiced. Aspiration is an optional feature of the voiceless stops. The phonological specification is mirrored in the voicing profiles in Figure 9, which shows the effect of the factor phone identity. The voiced stops population is neatly separated from the voiceless stops, with the glottal stop in a somewhat ambiguous position.

By far the strongest factor influencing the voicing profiles of stop closures in German is the preceding context phone. The effect is displayed in Figure 10, where the context phones are

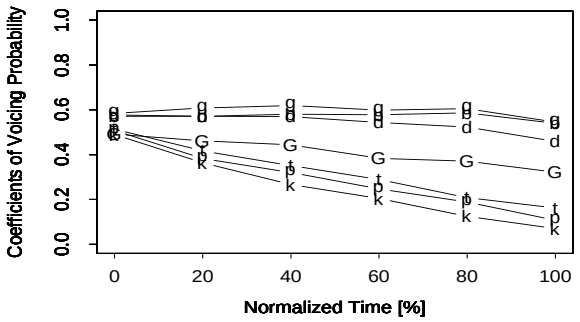

Figure 9. German voicing profile: effects of phone identity. Phonologically voiced and voiceless stops are also separated along the acoustic voicing dimension. "G" = glottal stop.

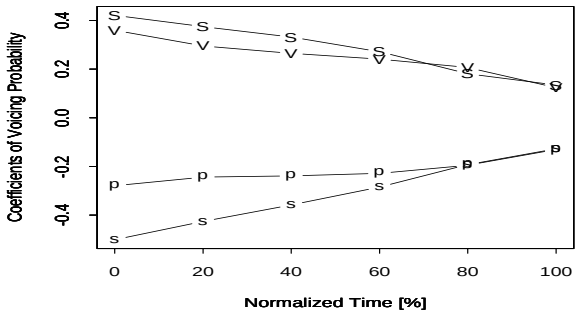

Figure 10. Effects of the preceding phone in German. Preceding voiceless obstruents have a strong devoicing effect, sonorant and vocalic left contexts cause the closure to be voiced. "V" = vowel, "S" = sonorant, "p" = voiceless stop, "s" = voiceless fricative.

categorized into vowels, sonorants, voiceless stops, and voiceless fricatives. As one would predict, preceding voiceless obstruents have a strong devoicing effect on the stop closure, whereas a voiced preceding context phone causes the closure to be voiced almost throughout.

4. DISCUSSION

We were interested in discrepancies between the phonological voicing status of a speech sound and its actual phonetic realization in connected speech. For instance, stop consonants in the five languages under investigation can be described using different combinations of the features of voicing and aspiration. Yet, we observe that, despite the difference in phonological specification, Mandarin voiceless unaspirated stop closures show voicing profiles similar to the voiced unaspirated stop closures in German. Similarly, the voiceless aspirated stops of Mandarin pattern with the voiceless (aspirated or unaspirated) stops in German, Spanish, and Italian. Hindi presents the most complex stop consonant system among the five languages in that it has a two-way phonological distinction between voiced/voiceless and aspirated/unaspirated stops. In our data, Hindi stops can be ranked along this scale of decreasing probability of voicing, e.g., /g, gh/ $\gg$ /k/ /kh/.

Our research shows that the phonological specification of voicing, represented as a binary distinction of [+voice] or [-voice] over the domain of the entire speech sound, is often insufficient to differentiate the stop series in a given language. It also obscures similarities or parallel patterns across languages. VOT measurements might provide a better classification of stops.

However, note that in both Mandarin and German the two populations of stops are differentiated by the patterns of sustained voicing in the stop closure duration. Voicing profiles, as suggested in our study, allow us to describe the dynamic changes of the voicing status of speech sounds, here stops, as a function of (normalized) time. In the conventional usage of VOT, being voiced is expressed as a negative VOT value counting backward from the time of the stop release. Since voicing typically ceases before the burst in all stops of Mandarin and German, the more voiced and the less voiced populations in these languages cannot be differenciated by a negative VOT value alone.

5. CONCLUSION

We presented results from a study of the voicing profiles of consonants in five languages, Mandarin Chinese, German, Hindi, Mexican Spanish, and Italian. We examined the factors that cause variations in voicing and trained statistical models that predict the probability of voicing of each analysis frame of a speech sound from a variety of factors, in particular phone identity, context phones, positional and prosodic factors.

This paper has focussed on the voicing properties of stop closures in these languages. Details on other speech sounds, especially voicing probability models for all consonant types in German and Mandarin, have been presented elsewhere [13].

REFERENCES

[1] Bell-Berti, F. 1975. Control of pharyngeal cavity size of English voiced and voiceless stops. *J. Acoust. Soc. Am.*, 57, 456–461.

[2] Chomsky, N., and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

[3] Dixit, R. P., and Brown, W. S. 1985. Peak magnitudes of oral air flow during Hindi stops (plosives and affricates). *J. Phonetics*, 13, 219–234.

[4] Keating, P. 1984. Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286–319.

[5] The Kiel Corpus of Read Speech, vol. 1. 1994. Publ. by IPDS, Univ. Kiel, Germany. CDROM.

[6] Ladefoged, P., and Maddieson, I. 1996. *The Sounds of the World's Languages*. Cambridge: Blackwell.

[7] Liberman, A., Delattre, P., and Cooper, F. 1958. The role of selected stimulus variables in the perception of voiced and voiceless stops in initial position. *Language and Speech*, 1, 153–167.

[8] Lisker, L., and Abramson, A. S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.

[9] Löfqvist, A., and Gracco, V. L. 1994. Tongue body kinematics in velar stop production: influences of consonant voicing and vowel context. *Phonetica*, 51, 52–67.

[10] Ohala, J., and Riordan, C. 1979. Passive vocal tract enlargement during voiced stops. In *Speech Communication Papers presented at the 97th ASA Meeting (New York)*, 89–92.

[11] Perkell, J. S., Holmberg, E. B., and Hillman, R. E. 1991. A system for signal processing and data extraction from aerodynamic, acoustic and electroglottographic signals in the study of voice production. *J. Acoust. Soc. Am.*, 89, 1777–1781.

[12] Poon, P. G., and Mateer, C. A. 1985. A study of VOT in Nepali stop consonants. *Phonetica*, 42, 39–49.

[13] Shih, C., and Möbius, B. 1998. Contextual effects on voicing profiles of German and Mandarin consonants. In *Proc. 3rd Int. Workshop on Speech Synthesis (Jenolan Caves, Australia)*, 81–86.

[14] Shimizu, K. 1990. A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages. PhD thesis, Univ. Edinburgh.

[15] Slis, I. H., and Cohen, A. 1969. On the complex regulating the voiced-voiceless distinction I. *Language and Speech*, 12, 80–102.

[16] Stevens, K. N. 1988. Modes of vocal fold vibration based on a two-section model. In O. Fujimura (ed.), *Vocal Physiology: Voice Production, Mechanisms and Functions*, New York: Raven, 357–367.