ACOUSTIC CUES TO MANNER OF ARTICULATION OF OBSTRUENTS IN MARATHI

Vaishali Patil, Preeti Rao
Department of Electrical Engineering
Indian Institute of Technology, Bombay
Mumbai, INDIA 400076
Email: {vvpatil, prao}@ee.iitb.ac.in

Abstract

Acoustic cues are investigated for manner classification of unvoiced obstruents in Marathi. A statistical analysis of three temporal parameters extracted from the signal in the vicinity of the noise burst indicates the potential of such features in the three-way classification of unvoiced obstruents into fricative, stop and affricate manners. While unaspirated obstruents are accurately identified, the temporal features are ineffective in separating aspirated stops from affricates.

Keywords: *manner of articulation, obstruents, frication duration, rise time, rate of rise*

I] Introduction

Knowledge-based approaches to speech recognition use explicit knowledge about speech in the recognition process. This can be justified to be efficient, with the emphasis on processing speech rather than signals in general. Research on discovering acoustic attributes of linguistic features forms the backbone of knowledge-based systems. Given this perspective, Indian languages have not been studied well enough. The present investigation focuses on exploring acoustic-phonetic features to distinguish obstruent consonants of Marathi on the basis of manner of articulation. Obstruents are typically classified into the manner classes, *affricate*, *stop* and *fricative*. Stops are characterized by a complete closure located in the vocal tract, followed by an abrupt release. Fricatives are articulated by a narrow constriction in the vocal tract, giving rise to a steady frication noise. Affricates lie between stops and fricatives in acoustic properties. While they are produced by complete constriction, the release is accompanied by a distinct frication noise. A number of studies have investigated temporal properties extracted from the speech signal in the vicinity of the release/frication burst for cues to the three manner classes [1,2,3,4].

In an early study, Gerstman [5] used subjective perception tests on manipulated speech signals to demonstrate the importance of frication duration in identifying stops, affricates and fricatives. Howell and Rosen [2] measured distinct differences in rise time of the frication between the voiceless affricate (‘tʃ’) and fricative (‘ʃ’) in British English. A cross language study of Mandarin, Czech and German was carried out by Shinn [6] on the stop-affricate-fricative distinction from temporal measurements of rise time, rate of rise and frication duration. Rise time and frication duration were found to provide relatively effective cues to manner. Kluender and Walsh [7] found that rise time was not an adequate cue in a perception test involving voiceless affricates and fricatives. More recently, Hoelterhoff and Reed [4] showed that total consonant duration was very effective in distinguishing stops from affricates and fricatives in word medial and final positions in German.

In the present work, we investigate the effectiveness of the acoustic parameters of frication duration, rise time and rate of rise in the identification of manner of articulation of Marathi unvoiced obstruents. Marathi (unlike English, but similar to several other Indian languages) is characterized by the presence of both unaspirated and aspirated unvoiced stops and affricates. Further, the stops and affricates are traditionally categorised by place of articulation rather than manner even though the places of articulation are in close proximity in some cases.

II] Database

Marathi is one of the prominent Indian languages with over 70 million native speakers. The Marathi phone set includes 6-affricates (3-unvoiced and 3-voiced), 16-stops (8-unvoiced and 8-voiced) and 3-unvoiced fricatives. The current task concentrates on the class of unvoiced obstruents. This set comprises of 3 unvoiced affricates (‘ts’, ‘tʃ’, ‘tʃ^{h}’), 8 unvoiced stops (‘p’, ‘ʈ, ‘t’, ‘k’, ‘p^{h}’, ‘ʈ^{h}’, ‘t^{h}’, ‘k^{h}’) and 3 unvoiced fricatives (‘s’, ‘ʃ’, ‘h’).

A word list was prepared that included two words of each consonant ‘C’ in word initial position along with each of the eight vowels resulting in 16 words per consonant. The words are embedded in two sentences (one statement and one question) to be recorded by each of 4 speakers. Thus for each consonant we have 32 tokens (2 words X 2-sentences/utterances X 8-vowel context) per speaker. The database thus includes a total of 384 tokens of affricates, 1024 tokens of stops, and 384 tokens of fricatives. The utterance of stop ‘p^{h}’, by all the speakers failed to exhibit the expected closure which was evident on observing the waveforms, and was excluded from the experiments, as was the glottal fricative ‘h’ due to semivowel features. Thus we have now 1024 tokens of unaspirated obstruents and 512 tokens of aspirated obstruents for the feature evaluation.

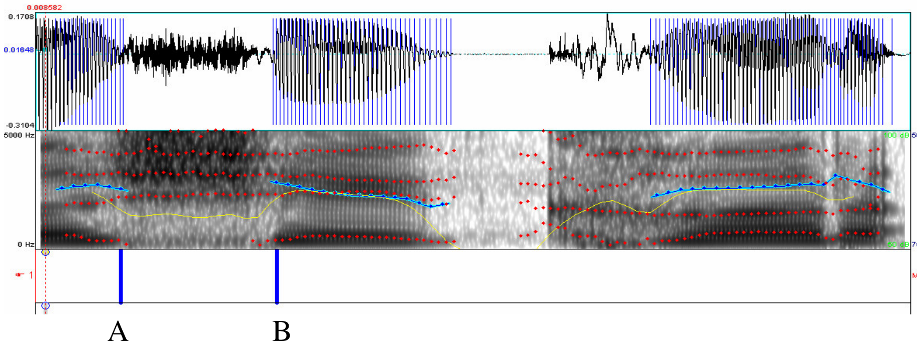

Fig. 1. Labeling of fricative (‘ʃ’), from word ‘ʃɛk^{h}ər’ (shekhar) of female speaker.

Next, segmentation and labeling are achieved manually using PRAAT [8] which facilitates simultaneous listening and observation of waveforms and spectrograms. The instants corresponding to the events of i) onset of noise (A) and ii) offset of noise (B) are marked in the initial obstruent of the extracted words. An example of manual labeling of the fricative ‘ʃ’ occurring in the word ‘ʃɛk^{h}ər’ (‘shekhar’) from the recording of a female speaker is shown in Fig 1. In the process of manual labeling, the time instants corresponding to the events marked are recorded in text file associated with each token.

III] Extraction and evaluation of temporal features

The temporal parameters considered for manner classification are i) frication (noise) duration, ii) rise time (RT), and iii) rate-of-rise (ROR). The parameters are extracted from each utterance by the analysis of speech samples in the neighborhood of the manually marked speech events corresponding to the noise burst. We do not consider total consonant duration due to the word-initial position used in our database, making the measurement of closure duration ambiguous. The evaluation of the parameters is carried out by computation of their statistical distribution parameters and depicted by box-and whiskers diagrams.

Frication duration

Frication duration is measured as the difference between the time instants marked by labels ‘A’ and ‘B’ in Fig. 1. The distribution of frication duration across unaspirated affricates, stops and fricatives is shown in Fig 2. We see that the mean frication duration values of the three classes are well separated and are respectively reflected in the F ratio and p values as F (2,1021) = 2965.04 (p=0). Affricate frication duration is shorter than that of fricatives but longer than that of stops. While frication duration can efficiently differentiate among the three unaspirated obstruent classes, there is a large overlap in the distributions of aspirated affricates and stops as seen in Fig 3.

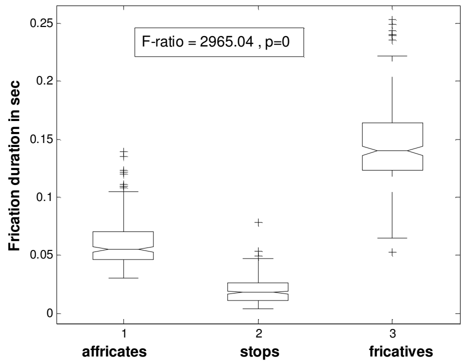

Fig. 2. Box-whiskers diagram of frication duration across unaspirated affricates, unaspirated stops and fricatives.

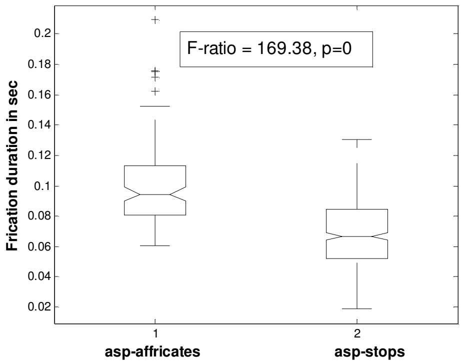

Fig. 3. Box-whiskers diagram of frication duration across aspirated affricates and stops

Rise time

Rise time is defined as the time from the onset of noise to the maximum amplitude in the noise (frication) duration. The total frication duration is divided into equal duration frames and average RMS energy is calculated for each frame in the frequency band of 2-8 kHz. Rise time is measured as the time interval between the frication onset and the instant of maxima in the average RMS energy envelope. Rise time is the least for stops corresponding to their shorter noise (frication) duration, and maximum for fricatives within the category of unaspirated obstruents. The F ratio and p values corresponding to the distribution of rise time are shown in Fig 4. Similar to frication duration, the distributions of rise time show large overlap for the two aspirated obstruent classes as seen in Fig. 5.

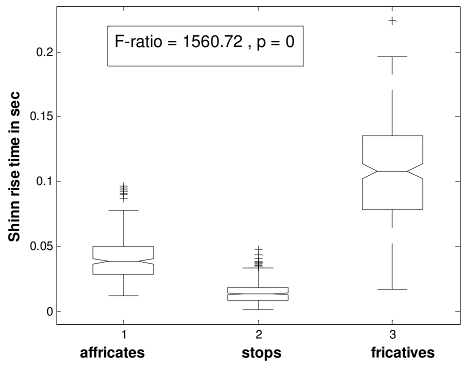

Fig. 4. Box-whiskers diagram of rise time across unaspirated affricates, unaspirated stops and fricatives

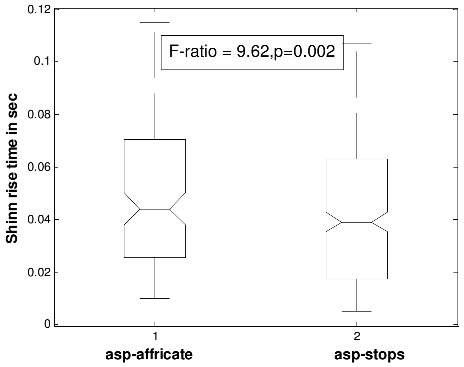

Fig. 5. ANOVA distribution of rise time across aspirated affricates and aspirated stops

Rate-of Rise (ROR)

The rate-of-rise (ROR) is defined as the derivative of the log rms energy with respect to time, in distinction of unvoiced plosives (stops and affricates) and fricatives. The log-rms energy is calculated at intervals of 1ms (c=1ms) using a standard Hamming window of 24 ms. (N=24ms). Eq. (1) gives the log-rms energy for the nth frame.

$$E_n = 10\log[(1/N) \sum_{i=cn}^{cn+N-1} [y(i)w(i-cn)]^2] \tag{1}$$

where w(i) is the standard hamming window, and y(i) is the preemphasized speech token.
The rate-of-rise is evaluated as the change in log-rms energy between the current window frame and previous window frame per unit time. It is defined as

$$ROR_n = (E_n - E_{n-1})/\Delta t \qquad (2)$$

where, Δt is the separation in ms, between two adjacent frames (1ms).
The ROR is then the value of the first peak (considering a 5-point neighborhood) in the ROR contour. The cases where no prominent peak is identified, ROR is assigned a value of '0'. Fig. 6 gives the distribution of unaspirated affricates, unaspirated stops and fricatives. We observe significant overlap.

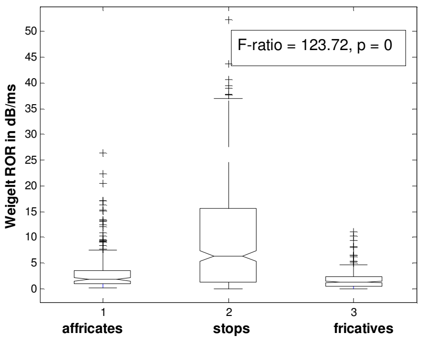

Fig. 6. Box-whiskers diagram of ROR across unaspirated affricates, unaspirated stops and fricatives

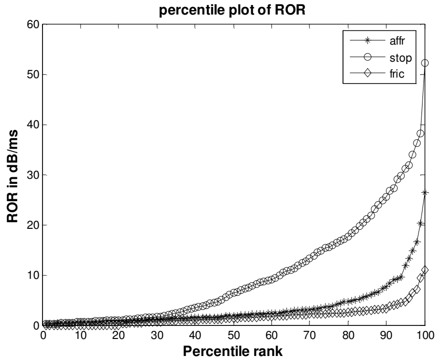

Fig. 7. ROR values in dB/ms versus percentile ranks separate for three unaspirated obstruent classes

This is consistent with the percentile plot of Fig. 7 (on the lines on Weigelt et al [6] for the corresponding obstruents in American English). The latter shows a clear threshold in ROR values that separates American English fricatives from the plosives, unlike that obtained in Fig. 7 obtained for the Marathi obstruents. We next investigate the possibility of observing ROR over a limited bandwidth for improved effectiveness in manner discrimination.

Liu [9] has used the rate-of-rise in different frequency bands to mark the onset of noise (burst) associated with stops and affricates. A wideband spectrogram is computed every 1 ms, using a 6 ms Hanning window. An ROR contour is obtained from the smooth energy waveform in each band (shown in Table 1) by taking overlapping first dB difference using 12 ms time separation, for every 1 ms. The value of first peak (considering 5 point neighborhood) in the ROR contour is assigned as the feature value in that band. Among the ROR evaluated in six bands, peaks in the 5th band are found to localize the noise onset with the best temporal accuracy. Hence ROR values in band 5 are selected as classification cues.

Band number	Frequency range
1	0.0 – 0.4 kHz
2	0.8 – 1.5 kHz
3	1.2 – 2.0 kHz
4	2.0 – 3.5 kHz
5	3.5 – 5.0 kHz
6	5.0 – 8.0 kHz

Table 1. Frequency bands for ROR computation

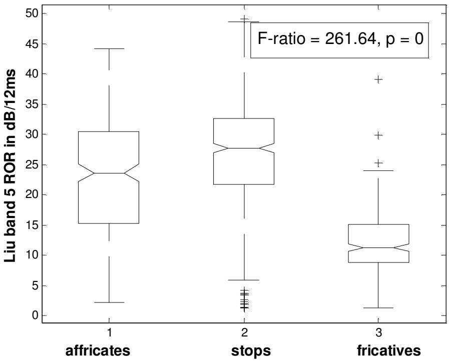

Fig. 8. Box-whiskers diagram of band 5 ROR across unaspirated affricates, unaspirated stops and fricatives.

Fig. 8 gives the distribution of band 5 ROR for the three classes of unaspirated obstruents. The distribution shows an improvement in the fricative-plosive distinction over Fig. 6, with increased overlap now between affricates and stops.

IV] Classification results and discussion

The classification experiments based on the three temporal features (frication duration, rise time and ROR in band 5 are carried using the hierarchical approach of decision trees [10]. The manner classification results of 4-fold cross validation on the entire set of unaspirated obstruents in shown Table 2. We see that fricatives are identified with high accuracy while a small degree of confusion prevails between stops and affricates.

Classified as / Actual token	Affricate	Stop	Fricative
Affricate	85.54%	8.59%	5.85%
Stop	5.07%	94.92%	0%
Fricative	4.68%	0%	95.31%
Average accuracy = 92.67%			

Table 2. Classification accuracy of unaspirated affricates, unaspirated stops and fricatives using temporal parameters of frication duration, rise time and band 5 ROR.

To summarize, the temporal features presented in the past literature, for manner of articulation detection of obstruents, are found to be effective in the separation of unvoiced, unaspirated obstruents of Marathi. The present experiments need to be extended to include obstruents in word medial and word final positions. Further, the invariance of the acoustic properties across speaking rate needs to be verified as the manner distinction relies prominently on temporal cues. The aspirated obstruents of Marathi, however, show a strong overlap in the temporal feature distributions indicating that more research is needed to accurately distinguish aspirated stops from affricates. A combination of place and manner cues may be more effective in separating the aspirated obstruents.

REFERENCES

1. Z. Mahmoodzade and M. Bijankhan, “Acoustic analysis of the Persian fricative-affricate contrast”, *Proc. ICPhS XVI*, pp. 921-924, Aug. 2007.
2. P. Howell and S. Rosen, “Production and perception of rise time in the voiceless affricate/fricative distinction”, *J. Acoust. Soc. Am.,* vol. 73, no. 3, pp. 976-984, Mar. 1983.
3. L. F. Weigelt, S. J. Sadoff, and J. D. Miller, “Plosive/fricative distinction: The voiceless case”, *J. Acoust. Soc. Am.,* vol. 87, no. 6, pp. 2729-2737, Jun. 1990.
4. J. Hoelterhoff and H. Reed, “Acoustic cues discriminating German obstruents in place and manner of articulation”, *J. Acoust. Soc. Am.,* vol. 121, no. 2, pp. 1142-1156, Feb. 2007.
5. L. Gerstman, “Noise duration as a cue for distinguishing among fricative, affricate, and stop consonants”, *J. Acoust. Soc. Am.,* vol. 28, no. 1, pp. 160, Jan. 1956.
6. P. C. Shinn, “A cross language investigation of the stop, affricate and fricative manners of articulation”, Ph.D. thesis, Brown University, May 1985.
7. K. R. Kluender and M. A. Walsh, “Amplitude rise time and perception of the voiceless affricate/fricative distinction”, *Perception Psychophysics,* vol. 51, no. 4, pp. 328-333, Apr. 1992.
8. P. Boersma and D. Weenink, “Praat: doing phonetics by computer (Version 4.3.01) [Computer program]”, Retrieved from http://www.praat.org/ , 2005.
9. S. A. Liu, “Landmark detection for distinctive feature-based speech recognition”, *J. Acoust. Soc. Am.,* vol. 100, no. 5, pp. 3417-3430, Nov. 1996.
10. Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques”, Morgan-Kaufman, 2006.