# SPEAKER DEPENDENT FEATURES IN STOPS AND AFFRICATES OF NATIVE URDU SPEAKERS

*MUHAMMAD JAMAL SHEIKH*

## ABSTRACT

The primary objective of this paper was to find phonetic cues that can be used for distinguishing between various speakers. The paper provides a brief review of the earlier researches done and some of the systems already implemented worldwide. Experiments, based on native Urdu speakers, were conducted for this purpose, and the observations were analyzed. The results inferred and the analysis done have lead to some phonetic cues that can be used for speaker identification and verification.

## 1. INRODUCTION

Human beings are able to understand speech and identify speakers from their voices because of the transitional characteristic of speech (Das, Molla & Ali). The utility of identifying a person from the characteristics of his/her voice is increasing with the growing use of speech interaction with computers. Automatic identification of a speaker not only has security and access control applications, but can also be used for speaker specific speech message retrieval and speaker labeling of conversations. In many such applications, it is necessary that a person can be reliably identified using short speech segments without regard to the text spoken (Reynolds, 1992).

## 2. LITERATURE REVIEW AND PROBLEM STATEMENT

Although investigations into the ability of humans to differentiate voices has a history of more than 50 years, research on speaker recognition by computers dates from mid-1960's. However, it took a number of years for that research to achieve commercialization. The earliest of those commercial systems applied speaker verification to door-access control. Most of those systems were designed to accept text-dependent input via microphone. Texas Instruments developed the most successful of the early algorithms. The "TI algorithm" is still used in some commercial products (Markowitz, Judith. 2000).

The number of applications of speaker verification is increasing steadily. They perform a broad spectrum of functions, including monitoring convicted felons, securing data and data networks, protecting buildings and other physical locations, monitoring time-and-attendance of employees, and securing transactions over the telephone (Markowitz, Judith. 2000).

In 1996, Illinois Department of Revenue (IDOR) implemented a speaker verification system for the security of data. This system had approximately 650 users and no casualties have been reported till now. BMC software started providing enterprise level support with speaker verification system in 1999. They have reported no difficulties with the system despite 200 users. Moreover, Monitoring services, Inc. provides electronic monitoring of criminal offenders with the speaker verification tracking system. Another company, Home Shopping Network began speaker verification on its numbers with more than 50,000 members enrolled. They report 95 to 97% verification on calls placed by enrolled members (Markowitz, Judith. 2000).

The 1990s have witnessed the flowering of commercial speaker recognition. Algorithms diversified to include hidden Markov models, Gaussian classifiers, various types of neural networks, and performance enhancements, such as anti-speaker modeling (Markowitz, Judith. 2000).

At present, most speaker identification systems use cepstral or linear prediction (LP) based features. However, the performance of these systems degrades significantly with the presence of noise in the

training and/or the testing speech (Trent & Reynolds, 1994).

The purpose of this paper is feature extraction. A number of features are to be selected on the basis of which a particular speaker may be distinguished from another one uniquely.

## 3. METHODOLOGY

The speech identification features are acquired by signal processing technique. Time dependent frequency analysis i.e. spectrogram is used for this purpose.

### 3.1 Speakers

The recordings were done by a group of native Urdu speakers by using each of the words in sentences of daily use. The speakers consisted of four. All of these speakers were young middle-aged people around 18 – 22 years old. The recordings were done keeping in mind the way that the speakers talk in their daily routine.

The major obstacle in the analysis of the speech signals is the environment while the recordings is done. The placement of the microphone, the intensity of the signal and the noise in the environment has resulted in a varying signal for different speakers.

### 3.2 Feature Extraction

For the purpose of extracting speaker dependent features, the place of articulation is not to be taken into account. The plosives or stops at any place have been divided into four categories with regard to their manner of articulation.

- Voiceless
- Voiced
- Aspirated
- Breathy

The plosives that were chosen for this analysis were all velars. For each of these stops, an Urdu word was selected.

### 3.3 Analysis of Plosives

All the stops have a closure time followed by a high amplitude burst. For velar stops, this burst consists of two high intensity regions in the spectrogram. The burst is followed by a relatively short voice onset time (VOT).

For the analysis purpose, the following data related to each recording was noted:

- Duration of Closure
- Duration of VOT
- Duration of preceding vowel
- Duration of following vowel
- Formant frequencies while going into the closure
- Formant frequencies when VOT starts
- Formant frequencies when going out of the stop

#### 3.3.1 Voiceless Stop

The voiceless stop uttered by each of the speaker had a short preceding vowel and a long following vowel.

Voiceless velar stop: | k |
| n ɪ k a ʈ̚ |
Translation: points

During the closed phase of voiceless consonants, the vocal folds are in a wide-open position. This wide-open position offers little or no resistance to the flow of air in the mouth and so the mouth pressure rises rapidly and becomes equal to sub-glottal pressure. Since the vocal folds are opened at the same time as the lips are closed, there is no vocal pulsing at all during the closed phase. Due to high pressure, upon release, there is a burst consisting of a transient, step like increase in the pressure of air. This is followed by damped oscillation at resonant frequencies determined by location of sound source in the vocal tract and the shape of the vocal tract (Picket, J.M. 1999, pp 115 - 123).

The duration of glottal airflow was also measured.

### 3.3.2 Voiced Stop

For the voiced stop, the word chosen had a short vowel both before and after the target consonant.

Voiced velar stop: | ɡ |
| ə ɡ ə r |
Translation: If

During the closed period of voiced stops, the vocal folds continue to open and close, emitting pulses of sub-glottal air into the closed oral cavity. With time, the air pressure in the mouth increases until it is high enough to stop phonation or until the release of the oral closure. The strength and duration of the release bursts of the voiced stops is lesser than that for voiceless stops (Picket, J.M. 1999, pp 115 - 123).

Due to these pulses of airflow emitted during voicing, the glottal airflow graph shows a vibratory pattern. The duration of the vibrating pattern was also noted.

### 3.3.3 Aspirated Stop

The aspirated stop used for sampling had a short vowel before it and a diphthong following it.

Aspirated velar stop: | $k^h$ |
| d̪ ɪ $k^h$ a o |
Translation: Show

Aspiration is produced by turbulence at the glottis. This turbulence causes a somewhat noisy VOT. This VOT is normally longer in duration than for voiceless stops.

### 3.3.4 Breathy Stop

The aspirated version of the voiced stops is called breathy stop. The word selected for this analysis was a syllable initial consonant followed by a short vowel.

Breathy velar stop: | $ɡ^h$ |
| $g^h$ ə r |
Translation: House

Aspiration in the VOT of voiced stops also causes a noisy pattern. The voice onset time for breathy stops is larger than voiced stops. However, this VOT is less than that of aspirated plosives.

## 3.4 Analysis of Affricates

The affricates have a closure in the first part of the phoneme with a dental burst at the end of the closure. The burst is followed by a fricative part.

For the analysis purpose, the following data related to each affricate was recorded:

- Duration of Closure
- Duration of friction
- Duration of preceding vowel
- Duration of following vowel
- Formant frequencies while going into the closure
- Formant frequencies when friction starts
- Formant frequencies when going out of the affricate
- Glottal stop during the utterance
- Formant intensities during friction

### 3.4.1 Voiceless Affricate

Two words were used for the analysis of affricates, one each for voiceless and voiced affricates. The voiceless affricate had a short vowel before it and a long vowel afterwards.

Voiceless affricate: | tʃ |
| b ə tʃ.tʃ e |
Translation: Children

### 3.4.2 Voiceless Affricate

For experimental purpose, the voiced affricate selected also had a preceding short vowel and a following long vowel.

Voiced affricate: | dʒ |
| ɪ dʒ a z ə t̪ |
Translation: Permission

## 4. RESULTS

The observations that were seen from the above experiments were noted. For each word uttered by a speaker, 10 samples were noted.

The statistical details of these observations are shown in the results:

| \| nikat \| | 709 | | 859 | | 1233 | | 1283 | |
|---|---|---|---|---|---|---|---|---|
| | Avg | stddev | Avg | stddev | Avg | stddev | Avg | stddev |
| Closure Duration | 0.076 | 0.009 | 0.063 | 0.006 | 0.056 | 0.005 | 0.061 | 0.007 |
| Vot | 0.019 | 0.005 | 0.026 | 0.003 | 0.044 | 0.007 | 0.026 | 0.003 |
| Preceding vowel | 0.047 | 0.004 | 0.035 | 0.003 | 0.047 | 0.008 | 0.041 | 0.009 |
| Following vowel | 0.159 | 0.005 | 0.124 | 0.009 | 0.140 | 0.008 | 0.153 | 0.009 |
| f4 / f3 (in) | 1.461 | 0.028 | 2.108 | 2.244 | 1.545 | 0.091 | 1.512 | 0.102 |
| f4 / f3 (out) | 1.404 | 0.029 | 1.468 | 0.158 | 1.350 | 0.034 | 1.453 | 0.040 |
| Closure-Vot | 0.057 | 0.008 | 0.036 | 0.006 | 0.012 | 0.011 | 0.036 | 0.009 |
| following-preceding vowel | 0.112 | 0.004 | 0.089 | 0.009 | 0.092 | 0.012 | 0.113 | 0.008 |
| closure/vot | 4.170 | 0.929 | 2.396 | 0.347 | 1.317 | 0.316 | 2.444 | 0.549 |
| vot/closure | 0.253 | 0.066 | 0.425 | 0.059 | 0.798 | 0.184 | 0.427 | 0.090 |
| preceding/following vowel | 0.297 | 0.021 | 0.285 | 0.023 | 0.340 | 0.063 | 0.266 | 0.050 |
| following/preceding vowel | 3.381 | 0.244 | 3.533 | 0.298 | 3.027 | 0.540 | 3.872 | 0.640 |

| \| dɪk$^{h}$ao \| | 709 | | 859 | | 1233 | | 1283 | |
|---|---|---|---|---|---|---|---|---|
| | Avg | stddev | Avg | stddev | Avg | stddev | Avg | stddev |
| Closure Duration | 0.057 | 0.009 | 0.037 | 0.007 | 0.058 | 0.007 | 0.046 | 0.007 |
| Aspirated vot | 0.064 | 0.004 | 0.056 | 0.008 | 0.079 | 0.012 | 0.062 | 0.008 |
| Preceding vowel | 0.056 | 0.007 | 0.039 | 0.006 | 0.069 | 0.012 | 0.045 | 0.005 |
| Following vowel | 0.157 | 0.006 | 0.146 | 0.017 | 0.190 | 0.017 | 0.191 | 0.041 |
| f4 / f3 (in) | 1.406 | 0.091 | 1.581 | 0.044 | 1.442 | 0.149 | 1.407 | 0.222 |
| f4 / f3 (out) | 1.307 | 0.031 | 1.388 | 0.055 | 1.395 | 0.068 | 1.322 | 0.086 |
| Vot-closure | 0.006 | 0.011 | 0.019 | 0.010 | 0.021 | 0.016 | 0.016 | 0.009 |
| following-preceding vowel | 0.101 | 0.009 | 0.107 | 0.018 | 0.122 | 0.023 | 0.146 | 0.039 |
| closure/vot | 0.907 | 0.170 | 0.679 | 0.144 | 0.746 | 0.149 | 0.752 | 0.148 |
| vot/closure | 1.144 | 0.245 | 1.537 | 0.341 | 1.397 | 0.318 | 1.371 | 0.238 |
| preceding/following vowel | 0.356 | 0.047 | 0.273 | 0.052 | 0.364 | 0.073 | 0.253 | 0.090 |
| following/preceding vowel | 2.860 | 0.417 | 3.785 | 0.680 | 2.845 | 0.578 | 4.229 | 0.926 |

| \| agar \| | 709 | | 859 | | 1233 | | 1283 | |
|---|---|---|---|---|---|---|---|---|
| | Avg | stddev | Avg | stddev | Avg | stddev | Avg | stddev |
| Closure Duration | 0.068 | 0.004 | 0.036 | 0.007 | 0.034 | 0.008 | 0.053 | 0.009 |
| Vot | 0.013 | 0.003 | 0.011 | 0.003 | 0.011 | 0.005 | 0.018 | 0.005 |
| Preceding vowel | 0.063 | 0.006 | 0.059 | 0.011 | 0.062 | 0.011 | 0.061 | 0.008 |
| Following vowel | 0.102 | 0.009 | 0.067 | 0.005 | 0.100 | 0.009 | 0.108 | 0.011 |
| f4 / f3 (in) | 1.439 | 0.052 | 1.575 | 0.131 | 1.472 | 0.044 | 1.481 | 0.121 |
| f4 / f3 (out) | 1.508 | 0.044 | 1.580 | 0.090 | 1.519 | 0.049 | 1.477 | 0.043 |
| Closure-Vot | 0.055 | 0.005 | 0.025 | 0.006 | 0.023 | 0.008 | 0.035 | 0.011 |
| following-preceding vowel | 0.039 | 0.012 | 0.008 | 0.010 | 0.038 | 0.019 | 0.046 | 0.012 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| closure/vot | 5.612 | 1.198 | 3.562 | 0.833 | 2.963 | 0.946 | 3.238 | 1.191 |
| vot/closure | 0.186 | 0.040 | 0.297 | 0.080 | 0.328 | 0.164 | 0.347 | 0.123 |
| preceding/following vowel | 0.626 | 0.091 | 0.875 | 0.148 | 0.633 | 0.153 | 0.573 | 0.090 |
| following/preceding vowel | 1.627 | 0.222 | 1.173 | 0.212 | 1.676 | 0.463 | 1.781 | 0.259 |

| | 709 | | 859 | | 1233 | | 1283 | |
|---|---|---|---|---|---|---|---|---|
| \| gʰar \| | Avg | stddev | Avg | stddev | Avg | stddev | Avg | stddev |
| Closure Duration | 0.088 | 0.009 | 0.046 | 0.007 | 0.047 | 0.007 | 0.054 | 0.011 |
| Aspirated vot | 0.012 | 0.002 | 0.028 | 0.006 | 0.056 | 0.012 | 0.035 | 0.008 |
| Preceding vowel | 0.083 | 0.007 | 0.058 | 0.005 | 0.078 | 0.013 | 0.087 | 0.012 |
| Following vowel | 0.074 | 0.006 | 0.161 | 0.255 | 0.093 | 0.014 | 0.101 | 0.013 |
| f4 / f3 (in) | 1.453 | 0.060 | 1.597 | 0.060 | 1.463 | 0.191 | 1.432 | 0.183 |
| f4 / f3 (out) | 1.473 | 0.028 | 1.370 | 0.094 | 1.457 | 0.044 | 1.220 | 0.176 |
| closure-vot | 0.076 | 0.009 | 0.018 | 0.010 | 0.008 | 0.012 | 0.020 | 0.015 |
| following-preceding vowel | -0.009 | 0.011 | 0.103 | 0.256 | 0.015 | 0.011 | 0.015 | 0.008 |
| closure/vot | 7.627 | 1.717 | 1.718 | 0.481 | 0.898 | 0.273 | 1.697 | 0.714 |
| vot/closure | 0.136 | 0.027 | 0.624 | 0.170 | 1.184 | 0.268 | 0.662 | 0.203 |
| preceding/following vowel | 1.130 | 0.154 | 0.702 | 0.262 | 0.840 | 0.113 | 0.858 | 0.077 |
| following/preceding vowel | 0.902 | 0.135 | 2.887 | 4.754 | 1.211 | 0.173 | 1.175 | 0.108 |

| | 709 | | 859 | | 1233 | | 1283 | |
|---|---|---|---|---|---|---|---|---|
| \| batʃ.tʃe \| | Avg | stddev | Avg | stddev | Avg | stddev | Avg | stddev |
| Closure Duration | 0.087 | 0.010 | 0.072 | 0.007 | 0.060 | 0.005 | 0.059 | 0.011 |
| Friction Duration | 0.046 | 0.005 | 0.048 | 0.002 | 0.064 | 0.005 | 0.064 | 0.010 |
| Complete Stop | 0.117 | 0.002 | 0.097 | 0.009 | 0.123 | 0.005 | 0.120 | 0.017 |
| Preceding vowel | 0.058 | 0.004 | 0.067 | 0.006 | 0.073 | 0.008 | 0.070 | 0.009 |
| Following vowel | 0.085 | 0.007 | 0.064 | 0.004 | 0.095 | 0.036 | 0.091 | 0.014 |
| f4 / f3 (in) | 1.469 | 0.048 | 1.289 | 0.106 | 1.342 | 0.041 | 1.314 | 0.091 |
| closure-friction duration | 0.041 | 0.014 | 0.024 | 0.008 | -0.004 | 0.008 | -0.005 | 0.014 |
| following-preceding vowel | 0.027 | 0.008 | -0.003 | 0.009 | 0.022 | 0.032 | 0.021 | 0.017 |
| closure/vot | 1.932 | 0.419 | 1.507 | 0.178 | 0.939 | 0.124 | 0.939 | 0.189 |
| vot/closure | 0.539 | 0.112 | 0.671 | 0.076 | 1.083 | 0.153 | 1.112 | 0.272 |
| preceding/following vowel | 0.683 | 0.075 | 1.053 | 0.143 | 0.815 | 0.154 | 0.786 | 0.157 |
| following/preceding vowel | 1.483 | 0.180 | 0.965 | 0.125 | 1.292 | 0.391 | 0.766 | 0.806 |

| | 709 | | 859 | | 1233 | | 1283 | |
|---|---|---|---|---|---|---|---|---|
| \| ɪdʒazat \| | Avg | stddev | Avg | Stddev | Avg | stddev | Avg | stddev |
| Closure Duration | 0.061 | 0.011 | 0.033 | 0.007 | 0.047 | 0.008 | 0.039 | 0.008 |
| Friction Duration | 0.025 | 0.001 | 0.024 | 0.003 | 0.023 | 0.008 | 0.023 | 0.004 |
| Complete Stop | | 0.009 | 0.020 | 0.006 | 0.042 | 0.023 | 0.026 | #DIV/0! |
| Preceding vowel | 0.085 | 0.005 | 0.074 | 0.007 | 0.064 | 0.009 | 0.046 | 0.010 |
| Following vowel | 0.128 | 0.008 | 0.109 | 0.009 | 0.127 | 0.014 | 0.144 | 0.010 |
| f4 / f3 (in) | 1.404 | 0.034 | 1.385 | 0.030 | 1.332 | 0.097 | 1.284 | 0.092 |
| closure-friction duration | 0.036 | 0.011 | 0.010 | 0.008 | 0.024 | 0.013 | 0.016 | 0.007 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| following-preceding vowel | 0.043 | 0.008 | 0.035 | 0.007 | 0.063 | 0.016 | 0.098 | 0.014 |
| closure/vot | 2.443 | 0.460 | 1.438 | 0.376 | 2.190 | 0.693 | 1.714 | 0.300 |
| vot/closure | 0.423 | 0.083 | 0.734 | 0.176 | 0.512 | 0.214 | 0.602 | 0.120 |
| preceding/following vowel | 0.663 | 0.050 | 0.681 | 0.050 | 0.509 | 0.097 | 0.322 | 0.069 |
| following/preceding vowel | 1.515 | 0.111 | 1.477 | 0.114 | 2.029 | 0.373 | 1.812 | 1.973 |

## 5. DISCUSSION

The statistical analysis of all the data generated throughout the experiment provides some relevant information about the cues that may be used for identifying different speakers.

### 5.1 Analysis of Plosives

#### 5.1.1 Closure Duration in relation to Voice Onset Time

The most significant cue encountered in the experiments was the closure duration of stops as compared to the voice onset time. The closure time and the VOT as well as their manipulation in different formulae give us enough information to categorize different speakers in various categories.

In the utterance of | nɪkat |, different speakers have different closure time and VOT. This range may be same for many speakers. However, the range of the ratio for the two durations is a distinguishable feature for speakers. The ratios for the four male speakers are 4.17, 2.39, 1.31 and 2.44 that are quite varying.

Moreover, the difference in the duration of these two values is also a valid measure. Different speakers can be easily categorized by the variation of the difference of stoppage time and VOT. Both these parameters can be used for the differentiation purpose in the utterance of both voiceless stops like | k | and voiced stops like | g |. The voiced plosives have a little smaller VOT, but still the ratio and difference between the closure and voice onset time are a good enough measure for our purpose.

For aspirated stop such as | $k^h$ |, the variability in the closure time and aspiration time is also significant. The aspiration for most of the speakers is more than the closure time, contrary to un-aspirated stops. However, the ratio and the difference of the two values is not that distinctly visible for different speakers.

Breathy stops are the most suitable of the stops for our experiment. The difference in closure time and the breathy noisy part vary a lot. Mostly, the noisy part is longer in duration than closures; however, some speakers prolong the closure to exceed the duration of aspiration.

According to details of the analysis, the ratio of the closure duration with the VOT is seemingly a better measure than their absolute difference, since the values for different speakers are not seen to overlap that much.

#### 5.1.2 Absolute closure and voice onset time

The absolute value of the closure duration and the VOT are also seen to vary in different speakers.

One of the male speakers tends to have a much longer VOT than other speakers. This difference was more evident in the voiceless stops, whether aspirated or un-aspirated.

On the other hand, the closure time was seen to vary more in voiced stops over a number of speakers. This difference may be due to a smaller VOT of voiced stops. The breathy stops also have a relatively lower degree of difference in their VOT than in the closure time.

In this regard, several special cases are also seen. For example, a few speakers may not be able to produce proper aspiration. This causes a very small VOT, which can be used as a very important cue for some people.

### 5.1.3 Length of the preceding and following vowel

Although there is a lot of variability among the utterances of one speaker, and a wide range of values are concluded as a result of adjacent vowel durations. However, some people are seen to prolong the earlier vowel a bit. Others elongate the ending vowel to some extent. The sum of the absolute duration both vowels is a strong cue for speaker identification.

### 5.1.4 Ratio of the preceding and following vowel

For some speakers, the duration of the earlier vowel and the latter vowel can be used as an identification cue. However, this ratio may overlap for many users. Although this ratio is not an overwhelming parameter for the sake of identifying people, but, the variability and the range of this ratio can result in a categorization of people.

### 5.1.5 Ratio of formant 4 and formant 3

The ratio of the fourth and the third formant is not a negligible fact. The formant frequencies can be noted when they are just coming out of the stop. The ratio specially becomes more and more significant as the pitch or fundamental frequency of the speakers vary. This ratio was seen to give very distinct results when the voices of male and female speakers were considered. It is not a useless factor when distinguishing two males, but the difference is not that distinctly visible?

### 5.1.6 Other Formants

The lower formants such as f1 and f2 vary a lot throughout the experiment for different speakers. The variation was seen also for the same speaker. According to the results found so far, no relation seems probable among the lower formants that can be used as a parameter for a speaker recognition model.

## 5.2 Analysis of Affricates

### 5.2.1 Closure and Friction Duration

The duration of closure and friction does not vary too much over a number of speakers, although, minute differences may be noted. Most of the speakers gave almost equal duration of friction, although the closure varied as sees in stops earlier.

### 5.2.2 Glottal Stop

An important feature found during the analysis is the glottal airflow allowed by the user, during the utterance of affricates.

The duration of the complete glottal closure is an important factor especially in voiceless affricates | tʃ |.

In voiced affricates, there is not actually a complete closure. The glottis allows airflow and pulses can be seen in the glottal flow analysis. The difference in various speakers is seen when some of them tend to extend the voicing over the fricative period. Others end the voicing at the end of the stop, not elongating it to the fricative part. Thus, a small complete closure is observed during this time.

### 5.2.3 Ratio of Closure and Friction Duration

The ratio of the closure and friction duration in a word varies significantly over a number of speakers.

Some speakers tend to have a longer closure part as compared to fricative part and others have an elongated friction. The difference between the closure and friction of affricates is (closure-friction) is even negative for some speakers thus categorizing them from others.

This difference is seen to be much more significant in voiceless affricates. For two speakers it was 0.939 and for the other two it was 1.507 and 1.931 respectively.

### 5.2.4 Preceding and Following Vowels

Regarding speaker identification, the duration of the preceding and final vowel is

not observed to be very helpful. Although the absolute duration can be used as a minor parameter, but no significant differences were seen.

**5.2.5 Friction and Following Vowel**

In some cases, the duration of the following vowel seemed to vary with the duration of the fricative part in the affricate.

Throughout the experiment, no exact relation was found between the two. However, the interrelation shows that this may be used as a parametric ratio for acoustic analysis of different speakers.

**5.2.6 The ratio of higher formants**

The ratio of f4 and f3 also gives us sufficient information for use as a parameter in our experiment, although very precise calculations are required in this case. Contrary to stops, the formants seem to be more distinguishing just when they are going into the closure.

**5.2.7 Intensities during friction**

The amplitude of all the formants during the friction part was measured for different speakers. Although, a minute variation was seen but no actual pattern or relation can be inferred from the data. Hence, supposedly, the intensity does not seem to play a crucial role in speaker identification.

## 6. CONCLUSION

In the complete research, some important cues were found and some seemingly important parameters were discarded. Although the results of the analysis may fail on larger sets of data, but discarding the variation in environment, the few factors that were seen to participate considerably in speaker identification are:

- Ratio of closure with VOT
- Complete glottal closure
- Length of adjacent vowels
- Ratio of preceding and following vowel
- Ratio of higher formants

The recordings and analysis done for this research was on a very small scale. For any significant result, much more data (statistics) are required. Nevertheless, this work can be used for creating initial parametric information for an immature speaker identification model.

Moreover, the analysis and the formulation of exact rules for the creation of a model that can achieve success will need hours of more detailed statistical analysis.

## 7. REFERENCES

Clark, John & Yallop, Collin. *An Introduction to Phonetics and Phonology.* 1990.

Das, Dipankar and Molla, M. Khademul Islam and Ali, M. Ganjer. “Endpoint Detection and Speaker Identification System for Isolated Utterance”. University of Rajshahi, Rajshahi-6205, Bangladesh

Markowitz, Judith. “Ieri, Oggi, Domani Speaker Recognition Yesterday, Today And Tomorrow”. Evanston, Il 60201 Usa. 2000.

Picket, J.M. *The acoustics of Speech Communication.* “Consonsants; voiced-unvoiced contrast“. Allyn & Bacon, 1999.

Reynolds, D. A. “A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification”. PhD Thesis, Georgia Institute of Technology, September, 1992

Trent, L. J., Rader, C. M., and Reynolds, D. A. “Using Higher Order Statistics to Increase the Noise Robustness of a Speaker Identification System”. Proceedings of the ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp 221-224, April 1994

# 8. APPENDIX

## Voice Onset Time

In studies of the perception of stop voicing, the delay between stop release and the onset of voicing is called the voice onset time (VOT). VOT is defined for stops as the time elapsing from the release of the occlusion to the beginning of the voicing. VOT is short for voiced stops, about 0 to 20 ms and long for voiceless stops, 30 ms or more. (Picket, J.M. 1999, p.125).

### Phonation and Glottal Air Flow

The periodic vibration of the vocal folds known as phonation provides the most important and acoustically efficient sound source in the vocal tract. The expiratory airflow from the lungs is modulated into a periodic vibratory cycle with regulated frequencies and intensities. The glottal low is plotted against time and this waveform has a harmonic spectrum with a slope of –12 dB. In normal speech this slope of the spectrum varies considerably depending on the phonatory settings. To some extent, this may depend on the speaker's speaking style and to some extent, it will reflect the speaker's personal voice quality and habitual phonatory settings. The slope of the spectrum is controlled largely by the rate of change of airflow during the phonatory cycle, usually its fall from peak to closure in the pitch pulse. (Clark & Yallop, 1990, pp 212)