A STUDY ON THE CONVERSATIONAL PATTERN OF BANGLA SPEECH OVER TELEPHONE CHANNEL

by

Md. Saifur Rahman

A dissertation submitted to the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, in partial fulfilment of the requirements for the degree of

Master of Science in Engineering

(Electrical and Electronic Engineering).

June 1989

Declaration

This is to certify that this work has been done by me and this has not been submitted elsewhere for the award of any degree or diploma or for publication.

(Md. Saifur Rahman)

Approval

Accepted as satisfactory for partial fulfilment of the requirements for the degree of Master of Science in Engineering (Electrical and Electronic Engineering).

Board of Examiners

1. Kazi Mohiuddin Ahmed, MScEngg, PhD, MIEB — Chairman and Supervisor — 8/7/89
   Associate Professor,
   Department of EEE,
   Bangladesh University of
   Engineering and Technology,
   Dhaka 1000.

2. M. A. Matin, ME, PhD — Member (Ex-officio) — M. A. Matin 8.7.1989
   Professor and Head,
   Department of EEE,
   Bangladesh University of
   Engineering and Technology,
   Dhaka 1000.

3. Md. Mujibur Rahman, MScEngg, PhD, FIEB — Member
   Professor and Dean,
   Faculty of EEE,
   Bangladesh University of
   Engineering and Technology,
   Dhaka 1000.

4. Saiful Islam, MScEng, PhD, CEng, MIEE, FIEB — Member — Saiful Islam 8.7.89
   Professor,
   Department of EEE,
   Bangladesh University of
   Engineering and Technology,
   Dhaka 1000.

5. Shamsuddin Ahmed, MScEngg, PhD, FIEB — Member (External) — 8.7.89
   Professor and Head,
   Department of EEE,
   ICTVTR,
   Gazipur.

Acknowledgement

The author owes the initiation to write the dissertation on "A Study on the Conversational Pattern of Bangla Speech over Telephone Channel", to the inspiring suggestions of his supervisor Dr. Kazi Mohiuddin Ahmed and expresses his great admiration for the keen interest taken and patience shown to him.

The author acknowledges the valuable suggestions and advice with this work received from Professor M. A. Matin, Head of the Department of Electrical and Electronic Engineering. The author is thankful to him for the great help of letting the author use the 4-channel tape recorder of the Acoustics Laboratory.

Acknowledgements are also made to Professor Md. Mujibur Rahman, Dean, Faculty of Electrical and Electronic Engineering for his allout support to this research.

The author is grateful to Mr. Mahmood Hasan Chowdhury and also to Mr. Syed Towfiqur Rahman of Computer Science and Engineering Department, BUET for their lively interset and spontaneous help in the preparation of the computer program.

The author thankfully acknowledges the valuable discussions occassionally made on the present problem with Mr. Farhad Fuad Islam, Asstt. Prof., Department of Electrical and Electronic Engineering.

Last but not the least, the author would like to register his thanks to all those people who have extended their helping hands in times of need.

Contents

Page

Page

Abstract

The on-off characteristics (presence and absence patterns) of conversational Bangla speech over telephone channel have been investigated in this work. For this purpose the electrical characteristics of Bangla speech pertinent to the problems of determining its presence on a telephone circuit have been studied after recording the speech of a number of speakers on an audio tape recorder. A simple method is suggested here to collect the statistical data of inter speech time-slots of conversational Bangla speech over telephone channel. A speech detector is designed to generate two distinct levels of output for presence or absence of speech. The output is then fed to the RS-232 C port of a personal computer to collect the on-off speech patterns.

The speech temporal parameters such as speech activity, average spurt time, average pause time, talkspurt rate, and the cumulative distribution functions of talkspurts and pauses of conversational Bangla speech are calculated using the generated data-base. The effects of fill-in and hangover on different speech temporal parameters are calculated. Proper mathematical models are suggested for the on-off speech characteristics.

Finally, the statistical behavior of inter speech time-slots in Bangla conversation over telephone channel is presented both from experimentally measured data and from the suggested mathematical models.

Chapter 1

Introduction

1.1 Importance of the analysis of Bangla speech over telephone channel

Bangla is the language of over 150 million people. Among them around 100 million live in Bangladesh, approximately 50 million in West Bengal, Tripura, Assam, Orissa and Bihar states of India, and a few million in Britain, USA and other countries [1]. There are about 4000 languages in the world. In reference [1] the relative positions of the languages from the point of view of number of speakers have been listed in the order of Chinese, English, Hindustani, Spanish, Russian, German, Japanese and Bangla. From this we can see that the relative position of Bangla is eighth. The importance of Bangla as a language can be understood from the fact that about 1600 newspapers, periodicals and journals are published in this language [1].

Bangla is now being extensively used in academic institutions, government offices, autonomous bodies and in almost all public affairs in Bangladesh. Different languages are now being extensively analyzed and used in various fields of scientific research. Bangla as a language also needs such an analysis in order to coup with the new advancements.

Every language has its individual accent (i.e. mode of pronunciation) and phonetics. In addition, every language shows its characteristic electrical properties when a speaker of one language talks over a telephone channel. The electrical behaviour such as frequency content, amplitude level etc. are very important for signal transmission and reception over telephone channel.

Long distance analog communication as well as digital communication are two-way. They use separate facilities for the two directions of transmission. Though the bandwidth requirement is high in digital communication, present trend is to replace the analog telephone networks with its digital counterpart due to relative advantages of the later. As the demand for the capacity of digital communication increases, the efficient use of communication channel bandwidth is one of the important problems in digital communications research. Of the two methods of bandwidth compression in digital coding of speech, one method is efficient source coding, and the other is speech interpolation. The speech interpolation method is utilized in two-way telephone communication (analog or digital) to ensure maximum utilization of the available channels [2].

In a normal telephone conversation each subscriber speaks less than half of the time of the total duration of talk. The remainder of the time is composed of listening, gaps between words and syllables, and pauses while the operator or subscriber leaves the line [6]. This on-off property of conversational speech is utilized to make efficient use of transmission channels by employing the method of speech interpolation. For this purpose it is necessary to find out the threshold level of the electrical signal to detect the presence and absence of speech over telephone channel.

Extensive research have been carried out on different electrical properties of English speech. The refernces [2] and [8] can be mentioned here as examples. On the otherhand, Bangla as a language has not yet been extensively analyzed for an electrical communication channel. The acoustical study of the vowel structure of Bangla language carried out by Pramanic [3] is an important work. P. K. Das [4] studied the information contents of messages in Bangla language and S. Z. Huque [1] studied statistically the time-amplitude pattern and the frequency bandwidth of Bangla alphabets in voice signals.

The purpose of this thesis is to study the on-off patterns (presence and absence) of conversational Bangla speech over telephone channel and to develope proper mathematical models for them. These on-off speech statistics are important for the design and performance analysis of speech interpolation systems for Bangla speech.

This analysis may further lead to a comparison of the on-off patterns of Bangla language with those of other languages and discover the noticeable differences, if any.

It is expected that on-off patterns of different languages would be different. Therefore, when a speech interpolation system has to be designed where people of different languages will talk over telephone channels of the system, due consideration should be given to the on-off patterns of all the languages. The on-off speech statistics of Bangla language may then be helpful in the design of the aforementioned speech interpolation system.

1.2 Objective of this work

The object of the study is to investigate the on-off characteristics of conversational Bangla speech over telephone channel and to develope proper mathematical models for them. For this purpose the following steps have been carried out:

1. The first step of the work is the preparation of the data-base (Bangla speech). Telephone conversations of different speakers are recorded on a tape recorder.

2. In the second step of the work an experimental set-up is prepared using electronic components to detect the presence and absence of speech.

3. In the third step the on-off patterns of conversational Bangla speech over telephone channels are generated and different speech temporal parameters are computed using computer facilities.

4. In the final step the statistical analysis of on-off patterns is performed which ultimately results in mathematical models for Bangla conversational speech over telephone channel.

1.3 This Dissertation

In chapter 2 a theoretical description of the method of speech interpolation is given. The design procedure of the speech detection unit and the technique of data collection are described in chapter 3. In chapter 4 the statistical analysis of Bangla speech data is presented and different speech temporal parameters are calculated. Based on the on-off data some simple mathematical models for cumulative distribution function of talkspurt and pauses and for other speech temporal parameters are developed and they are presented in chapter 5. The experimental results and discussions of this study are given in chapter 6. Finally, the thesis has been concluded with a conclusion and suggestions for further work in chapter 7. Appendices and references are enclosed at the end of this dissertation.

Chapter 2

The methods of speech interpolation

2.1 Introduction

Efficient use of communication channels is vital in rapidly expanding present day communication. Speech interpolation is an established technique for ensuring the efficient use of communication channel.

In this chapter we introduce methods of speech interpolation as used in present day communication.

2.2 Telephone transmission

Modern day telephone transmission could be analog or digital in nature. In a telephone network virtually all subscriber loops are implemented with a single pair of wires. The single pair provides for both directions of transmission. In contrast, wire-line transmission over longer distances, as between switching offices, usually involves two pairs of wires: one pair for each direction. Longer distance transmission requires amplification and often involves multiplexing. These operations are implemented most easily if the two-directions of transmission are isolated from each other [5]. This is why, digital communication and long distance analog communication are two-way.

The four-wire analog system uses two separate pairs of wires for transmit and receive directions. Figure 2.1 shows the schematic diagram of a four-wire analog system.

The use of four-wire transmission has a direct impact on the switching systems of the toll networks. Here toll networks are meant to be those networks which involve numerous switching offices with relatively long transmission links between them. Since most toll network circuits are four-wire, the switches are designed to separately connect both directions of transmission.

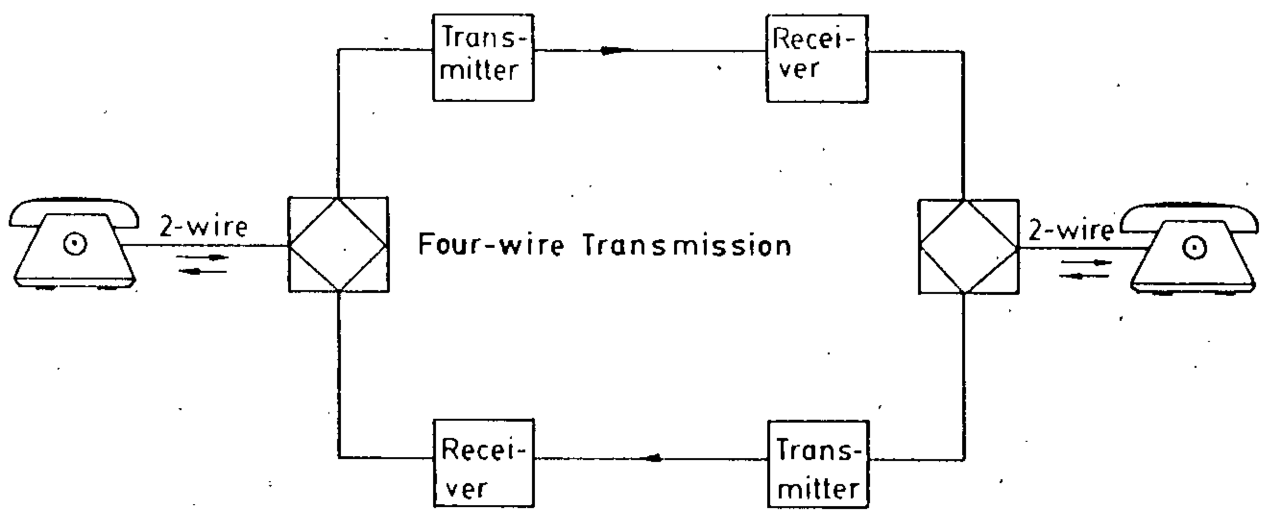


Figure 2.1 Four-wire analog system.

In digital transmission system the analog signals are sampled and PAM signals are obtained. These PAM signals are quantized and converted into digital signals by using pulse code modulation (PCM) technique for transmission. Like a four-wire analog system, separate physical pairs of wires are required for PCM transmit and receive directions. In each pair a particular time slot is alloted to a subscriber. Thus one-transmit and one-receive form one voice circuit. So PCM transmit and receive systems are regarded as the digital equivalent of four-wire analog system. Figure 2.2 shows a four-wire digital system.

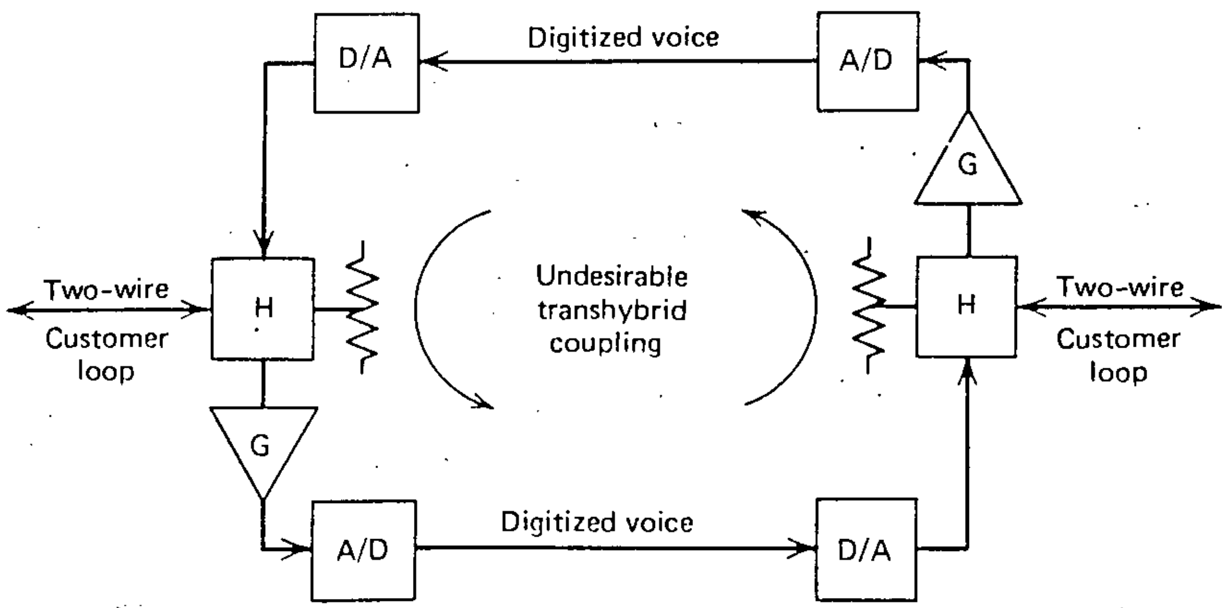


Figure 2.2 Four-wire digital switch circuit with two-wire analog interfaces.

2.3 Speech interpolation

Four-wire circuits use separate facilities for the two directions of transmission and during conversation each one-way channel is free (as already stated in chapter-1) about more than 50 percent of the total time. If one can efficiently use this silence portion of speech, the effective channel capacity can be greatly increased in circuit-switched voice transmission. The method of using this idle time in telephone calls, to interpolate additional talkers, is known as speech interpolation. This method is utilized in two-way telephone communication to ensure maximum utilization of the available channels [2]. Two types of speech interpolation techniques are in use. One is Time Assignment Speech Interpolation (TASI) and the other is Digital Speech Interpolation (DSI).

2.3.1 Time Assignment Speech Interpolation

Time Assignment Speech Interpolation (TASI), is a high-speed switching and transmission system which uses the idle time in telephone calls to interpolate additional talkers [6]. It is essentially a bank of voice operated switches which may disconnect a subscriber from a channel when he is not talking to permit a talking subscriber to use the channel [7]. This technique is used by AT&T to increase the capacity of analog circuits in transoceanic cables. TASI has been in service on transatlantic submarine cable channels since mid-1960.

Measurements on working transatlantic channels, show that a TASI speech detector with a sensitivity of -40 dbm is operated by speech from one talker on the average about 40 percent of the time the circuit is busy at the switchboard. Since long distance circuits use separate facilities for the two directions of transmission, each one-way channel is, on the average, free about 60 percent of the time.

In order to take advantage of this free time to interpolate additional conversations, a considerable group of channels must be available. An attempt to interpolate two independent conversations on a single channel would result in a large percentage of the speech being lost, since the probability of both talkers speaking at the same time is high. However, with a large group of channels serving a large group of talkers, the variations in demand become much smaller. Even with 74 talkers on 37 channels, the percentage of speech lost (freeze-out fraction) is reduced to a point where there is no noticeable effect on continuity of conversation.

The increase in channel capacity with TASI is illustrated in figure 2.3 [6], which gives the TASI advantage (ratio of switchboard positions, or trunks, to channels) for a range of activities and number of channels. A freeze-out fraction of 0.5 percent has been assumed for each curve, since this amount of speech loss has been found from tests to have a negligible effect on transmission quality.

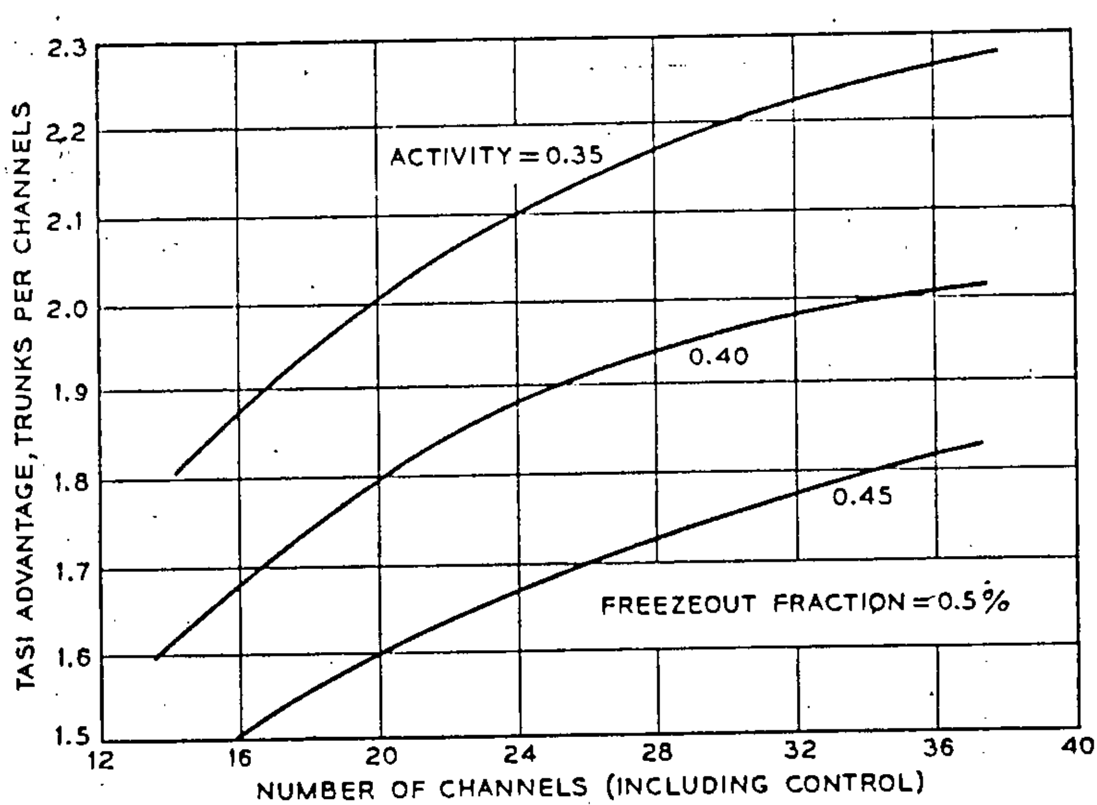


Figure 2.3 TASI advantage.

It will be noted in figure 2.3 that a TASI advantage of at least two can be obtained on a 37 channel group as long as the average activity is not significantly greater than 40 percent.

TASI is designed to use 36 channels for speech interpolation and one additional channel as a control channel for transmitting disconnect and error checking signals.

The first TASI system was put in service in June, 1960 on the transatlantic cable system between White Plains and London (TAT-1) and the second followed a few months later on the cable between New York and Paris (TAT-2) [6].

The Bell System had a plan to expand usage of TASI by applying it to domestic voice circuits beginning in 1983 [5].

2.3.2 TASI-operation

Figure 2.4 shows a block diagram of the TASI equipment for one direction of transmission; an independent TASI is used in the opposite direction of transmission [6].

Presence of speech on a trunk causes the speech detector to operate, initiating a request for a channel. The transmitting common control equipment selects an idle channel, if one exists, and assigns it to the requesting trunk.

Before the talker is connected to the channel, a "connect" signal is sent over the assigned channel specifying the trunk to be connected to that channel at the distant receiving terminal.

During the time required to connect talker and listener the initial part of the talker's speech is clipped. in order to minimize clipping the signaling time has been made as short as possible, 17 ms, consistent with reliable signaling and quiet switching. The signal information consists of a single burst of 4 tones out of a possible 14, ranging in frequency from 615 to 2419 Hz.

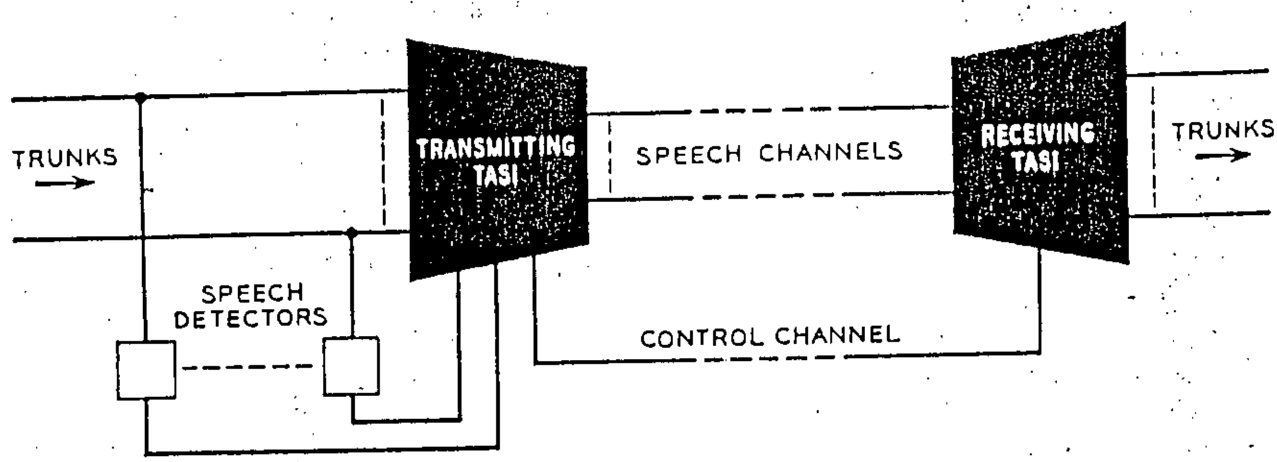


Figure 2.4 TASI equipment - one direction of transmission.

Once a talker is assigned a channel he does not lose the connection as long as he continues talking. When he stops talking he may still retain the connection unless he has to be disconnected to provide a channel for another talker.

A similar burst of 4 tones out of 15 (615 to 2501 Hz) is used to disconnect the talker and the listener, but this signal is sent over a separate control channel. During periods when no disconnect signals are being used, the same type of code signals are used to send information over the control channel as to the trunk-channel connections existing at the transmitting end. This connection-checking information overrides any earlier information and determines the connection made at the receiver. In addition, a comparision at the receiver between existing and overriding information is used to detect bad channels.

As shown in figure 2.3 the number of trunks which can be served by TASI depends upon the number of channels available. To prevent excessive speech loss when the connecting channels fail, TASI has been designed to automatically remove bad channels from service; trunks are then removed until the proper trunk-channel ratio is reached.

In addition to provisions for automatically reducing the number of connected trunks and channels, TASI contains audible and visible alarms to identify internal failures. In the event of a major failure in TASI the terminals automatically switch themselves out at both ends, reducing the number of connected trunks to the number of available channels. TASI is also switched out automatically if both the regular and alternate control channels fail. If only the regular control channel fails, the disconnect and error checking signals are automatically switched to the alternate control channel and TASI will continue to operate with only a momentary interruption.

When TASI is switched out, the voice frequency amplifiers associated with TASI are also switched out. The schematic relationship of these voice-frequency amplifiers and other transmission equipment is shown for one terminal in figure 2.5 along with typical operating level points [6]. The combination of VF amplifiers, TASI, and appropriate attenuation pads provides a zero-loss device and also provides optimum transmission levels to TASI.

The echo suppressor shown in figure 2.5 performs the usual function of preventing echoes, generated at points of impedance mismatch, from reaching a subscriber's ear and interfering with normal conversation. On TASI circuits, these suppressions must be of the receiving end split type at each end of the circuit to prevent the distant talkers echo from operating the speech detector. The location of a split suppressor is shown schematically in figure 2.5.

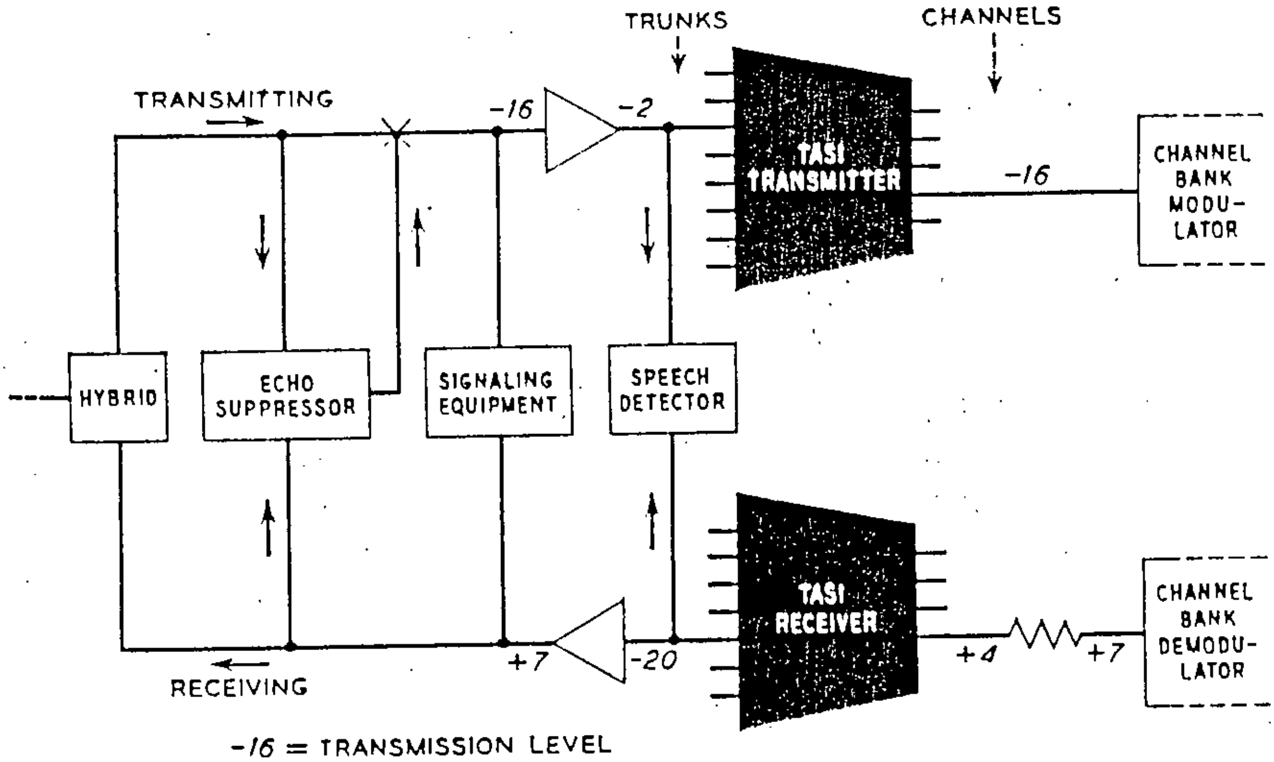


Figure 2.5 TASI and associated equipment.

2.3.3 Toll signaling and supervision

Because TASI is a time sharing device, there are problems involved in transmitting supervisory and dialing pulses. TASI can work satisfactorily with the present ringdown manual arrangement, but it is obvious that the usual method of continuous supervision by means of a steady tone during the idle time cannot be used. Likewise, dial pulses cannot compete for a TASI channel on the same basis as a talker, because TASI clipping would cause signaling errors. A burst signaling is required.

2.3.4 Engineering objectives for TASI

In order that TASI could operate over existing telephone facilities and would fit in with existing performance standards, certain engineering objectives were set up to guide the planning and development of TASI. They were considered as reasonable goals rather than rigid requirements. These objectives are listed in Table I [6].

Table I-Engineering Objectives for TASI

Capacity	At least 72 message trunks to be operated over 37, 3-kHz spaced cable channels. If the number of available channels is less than 37, the number of trunks to be provided by TASI will be less, as illustrated in figure 2.3. If the total busy hour speech activity is increased above about 40%, the maximum number of trunks to be provided will be less.
Speech quality	With the TASI system fully loaded as defined above, the degradation to speech quality due to TASI should not exceed about 1 db. When the number of talkers equals the number of available channels, the TASI degradation should be close to 0 db.

Signaling errors	On the average, during the busy hour, no more than 0.01% of the talkspurts transmitted should be lost because of signaling errors if the transmission medium meets the objectives noted in Table-II. Assuming that the average activity is 40%, this means about one talkspurt lost in thirty average 10-minute calls.
Reliability	The reliability objective is that the amount of time trunks are removed from service because of TASI failure shall be less than 0.1% of the total time.
Frequency response	The TASI transmitter and receiver connected back to back should pass a band of 200-3500 Hz. The average variation from flatness of all the channels should be within ± 0.5 db over this frequency range. In addition, the standard deviation of the variations from the average should not exceed 0.2 db.
Net loss	The net loss at 1000 Hz through the TASI equipment alone should be adjustable to within 0.15 db of 0 db and should stay within ± 0.15 db of the adjusted value for at least one month.
Circuit	The noise generated by TASI in the transmission path should not exceed about 12 dba as measured at the zero level points.

Crosstalk	To provide adequate crosstalk performance, an equal level coupling loss of 70 db should be obtained between talking paths in TASI. This applies to both near-end and far-end crosstalk.

2.3.5 Characteristics of channels for TASI

Because of the high-speed signaling used in TASI and because subscribers are switched rapidly between channels, the transmission requirements of the channels connecting TASI terminals are somewhat tighter than required for the usual telephone message service. The characteristics of importance to TASI are listed in Table II [6].

Table II-Required Transmission Characteristic of Connecting Channels for TASI

Minimum bandwidth	300 Hz -2900 Hz (10-db cutoff frequencies)
Flatness of band 565 to 2550 Hz	Difference between maximum and minimum loss should not exceed 2.5 db.
1-kHz net loss value of any channel	Not more than ±3 db from the nominal value.

Envelope delay distortion 565 to 2550 Hz	Not greater than 2 ms.
Flat delay	Maximum difference between channels should be be no more than 10-15 ms at 1000 Hz. (The control channel should be one of the fastest).
RMS noise	Without compandors, 38 dba at zero transmission level (38 dba 0). With companders, noise on line ahead of compandors should not exceed 51 dba 0. Difference in output noise between channels should not exceed 6 db.
Crosstalk	Equal level crosstalk loss on all channels should be at least 60 db.
Frequency stability (565 to 2550 Hz)	No frequency shifted more than 2 Hz.
Working levels for TASI (excluding pads or amplifiers outside of TASI)	Transmitting terminal input, -2 with respect to zero transmission level point. Receiving terminal input, +4 with respect to zero transmission level point.

2.3.6 Sources of speech clipping in TASI

The two major components of speech clipping in TASI are:

1. Signaling clipping, which is the time lost (17 ms) while a new connection is established;
2. Freeze-out, which is the time lost because no channels are available.

The length and frequency of these clips will vary from call-to-call due to loading, speech habits, or statistical chance.

In addition to signaling and freeze-out, some clipping is also caused by

(a) speech detector response time and threshold;

(b) disconnection delay caused by control channel crowding during heavy loading.

Inspite of these minor drawbacks, TASI has been operated successfully on submarine cable systems to provide approximately twice as many good quality message trunks as existed before TASI.

2.3.7 Digital Speech Interpolation

Recent use of the speech interpolation technique for digital voice signals is referred to as Digital Speech Interpolation (DSI). It is the digital circuit counterpart of TASI. In essence DSI involves: sensing speech activity , seizing a channel, digitally encoding and transmitting the utterences, and releasing the channel at the completion of each speech segment [5].

In digital transmission of speech a particular time slot is alloted to a subscriber in each channel. DSI utilizes that particular time slot of the channel when it is idle.

Owing to long holding times and only modest inactivity factors, voice traffic is most appropriately transmitted by way of the backbone circuit-.

switched network. On any particular route, however, the backbone transmission links can be implemented with DSI techniques to improve the circuit utilization. Thus the complexity of a DSI operation is added only where the cost is warranted.

When the number of active sources exceeds the number of channels in a DSI system, the usual operation is to block the overload and cause clipping of some speech syllables. More recently, however, a TASI system has been developed by STC Communications of Broomfield, Colorado that delays blocked syllables for a short period of time in anticipation that a channel will soon become available. The reconstructed speech from this process is of better quality than when conventional syllable clipping occurs. Thus greater levels of line utilization are possible when delay capabilities are incorporated into a DSI or TASI system.

In our study, we have performed the statistical analysis of the on-off patterns of conversational Bangla speech over telephone channel. The results of this study will be helpful in the design and performance analysis of any speech interpolation system for Bangla speech whether TASI or DSI.

In the next chapter the electrical characteristics of Bangla alphabet are tabulated and at the same time the technique of on-off data collection of conversational Bangla speech is presented.

Chapter 3

Conversational patterns of Bangla speech

3.1 Introduction

The theoretical aspects of speech interpolation systems have been described in the previous chapter. Our present aim is to carry out an analysis of the conversational Bangla speech over telephone channels to see whether and to what extent Bangla conversational speech may be interpolated.

This chapter deals with the electrical characteristics of Bangla alphabet and the characterisation of different speech temporal parameters of the conversational speech over telephone channels. In addition, the design procedure of the speech detection unit and the technique of measuring the durations of presence and absence of speech with the help of a personal computer are also described.

3.2 Characteristics of Bangla alphabet

Amplitude and duration of Bangla vowels and consonants uttered in isolation by male and female voices are studied by S.Z. Huque [1]. The results are given in tables III and IV in Appendix A.

Extensive study from the recorded data shows that both the amplitude and the duration of a letter vary from word to word. This variation is entirely dependent on the structure of the word and its mode of pronunciation. The distribution of amplitude and duration of a letter occuring in different words depends on the length of the word. It is also noted that for a closed monosyllabic word the amplitude of the first letter is very near to its amplitude uttered in isolation.

3.3 Characterisation of different speech temporal parameters

The on-off patterns of conversational speech are the patterns of its presence and absence over telephone channel. To obtain a correspondence to the presence or absence of speech, we define a talkspurt and a pause.

A talkspurt (or simply a spurt) is a time period which is judged by a listener to contain a sequence of speech sounds unbroken by a pause.

A pause is a time period which is judged by a listener to be a period of nontalking, other than one caused by a stop consonant, a slight hesitation or a short breath.

The function of the speech detector unit is to generate the talkspurt and pause patterns. Other than speech signal, noise occasionally operates the speech detector for short periods, and the resulting spurts should be discarded or thrownaway. The maximum time length of such a short period is considered as the throwaway time.

There are some small gaps caused by a stop consonant, a slight hesitation or a short breath within a continuous speech segment. These are called intersyllabic gaps. While processing speech data for analysis these gaps are filled in. The maximum duration of such a gap is considered as the fill-in time.

Most of the existing designs of speech detectors employ a slow release, or hangover, to bridge the intersyllabic gaps. Practically there is no other way but to employ a slow release to bridge those gaps. The time of slow release or delay is termed as hangover time. It is logical to choose this hangover time equal to the maximum duration of an intersyllabic gap.

On-off speech characteristics may be represented by probability distributions of talkspurt and pause durations, and by speech temporal parameters such as speech activity, average talkspurt and pause durations, and average talkspurt rate [2].

Speech activity represents the percent of time a party is active while talking over a telephone. This is the ratio of the actual time a person is talking to the total call duration.

Average talkspurt duration is obtained by summing the durations of all the talkspurts and dividing it by the total number of talkspurts in a given time.

Similarly, average silence or pause duration is obtained by summing the durations of all the pauses and dividing it by the total number of pauses in a given time.

The talkspurt rate is defined as the number of talkspurts per minute.

3.4 The design of the speech detection unit

The on-off patterns of conversational speech represents the patterns of its presence and absence over telephone channels. A speech detection unit has to be designed for the purpose of detecting the presence and absence of speech.

3.4.1 On-off patterns from different speech detectors

The task of detecting the presence or absence of speech is very difficult. Speech has a large dynamic range, and its levels frequently fall into the noise, even during segments audible to a listener. In addition, momentary interruptions due in part to stop consonants might cause a speech detector to indicate a silent interval whereas a listener would sense a continuing flow of speech [8]. These small gaps are called intersyllabic gaps and we should not use them for speech interpolation. Most existing designs of

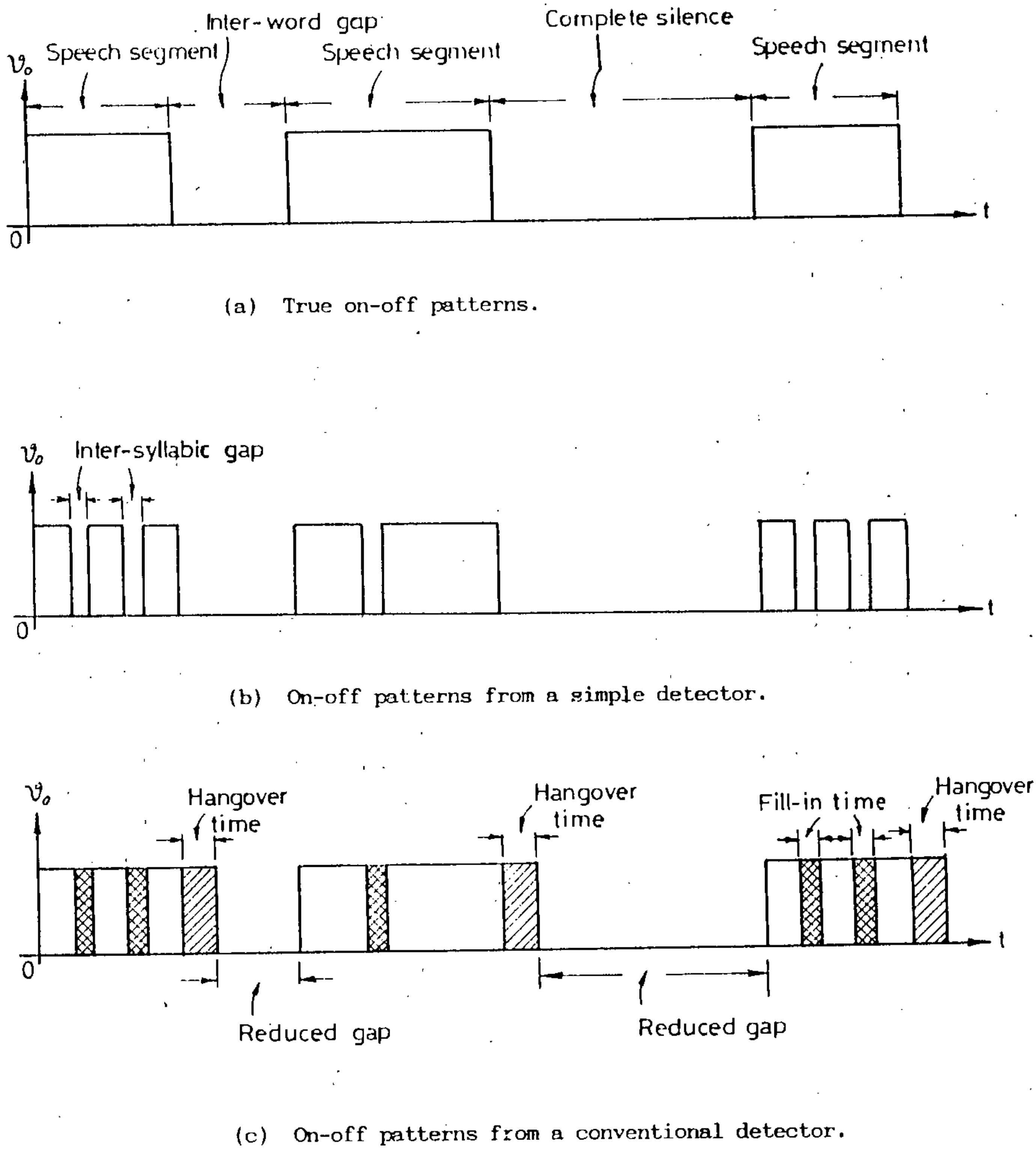


Figure 3.1 On-off patterns from different speech detectors.

speech detectors employ a slow release, or hangover to bridge such gaps, but an error equal to the hangover (delay) time is made everytime the person actually stops talking [8]. That is, the length of the time-segment when a person is not actually talking is produced less than what it is [Figure 3.1].

Therefore, if the analysis is performed with a detector having a slow release (delay), the result would be erroneous. In order to use a simple speech detector and at the same time to avoid some of the aforementioned shortcomings, a two-step detection technique is suggested here. Speech is first played through a speech detector, whose output is then processed by a computer program. It should be mentioned here that this is not the same detector as the one used in speech interpolation. It is the one which is used here for the purpose of analysis only.

3.4.2 Generation of speech data-base

To investigate the on-off characteristics of conversational Bangla speech we may use as the data-base two-way telephone conversations spoken in Bangla by a number of pairs of speakers. Telephone conversations of 20 speakers are recorded on a tape-recorder for each one-way channel. For ease of recording, dynamic microphone is placed close to the mouth-piece of the telephone hand-set and speech is recorded separately. Due consideration is given to ensure the amplitude level of the recorded signal to be the same as on the telephone circuit. Speech signals include both male and female voices and they are recorded from persons of different ages and of different professions while they are engaged in talking to their friends or relatives. The particulars of 10 male and 10 female speakers are listed in Appendix B. Every effort is taken to keep the conversations unbiased. For a better understanding of the experimental procedure, the block diagram of the speech detector that is used in this study is shown in figure 3.2.

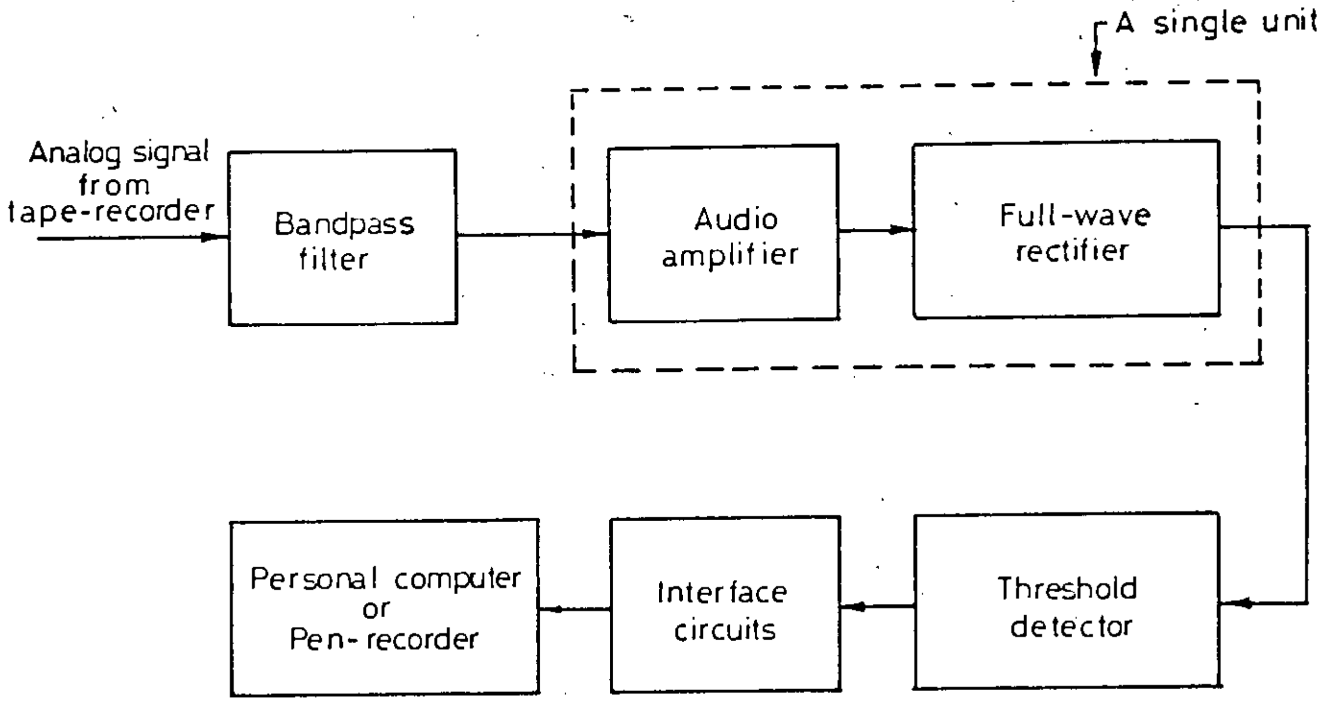
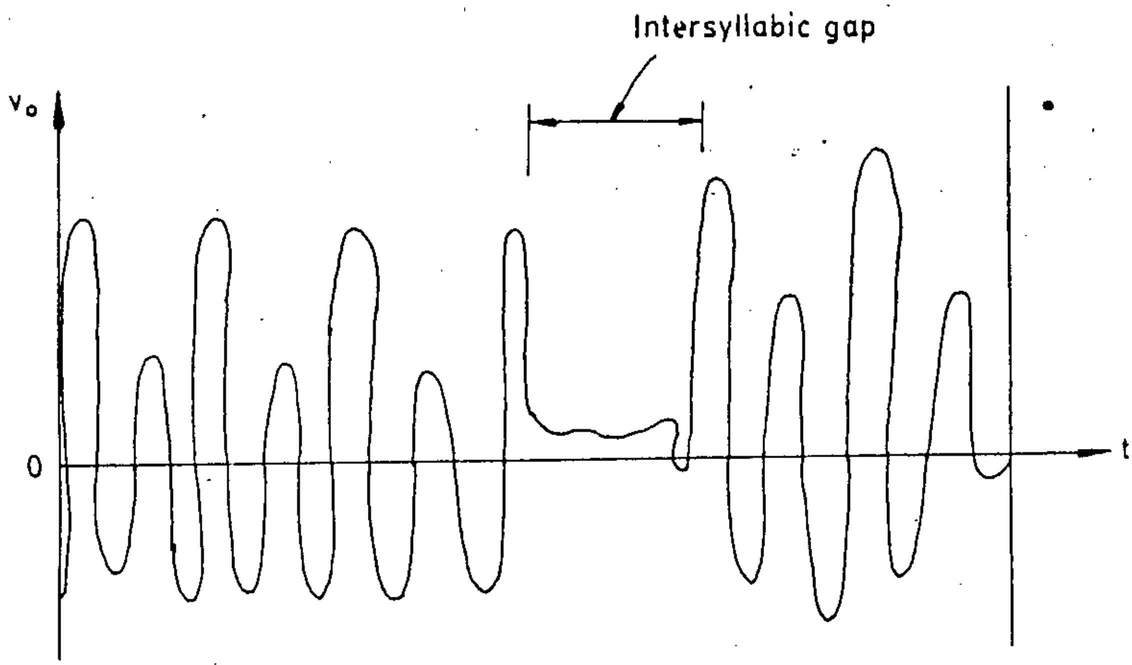


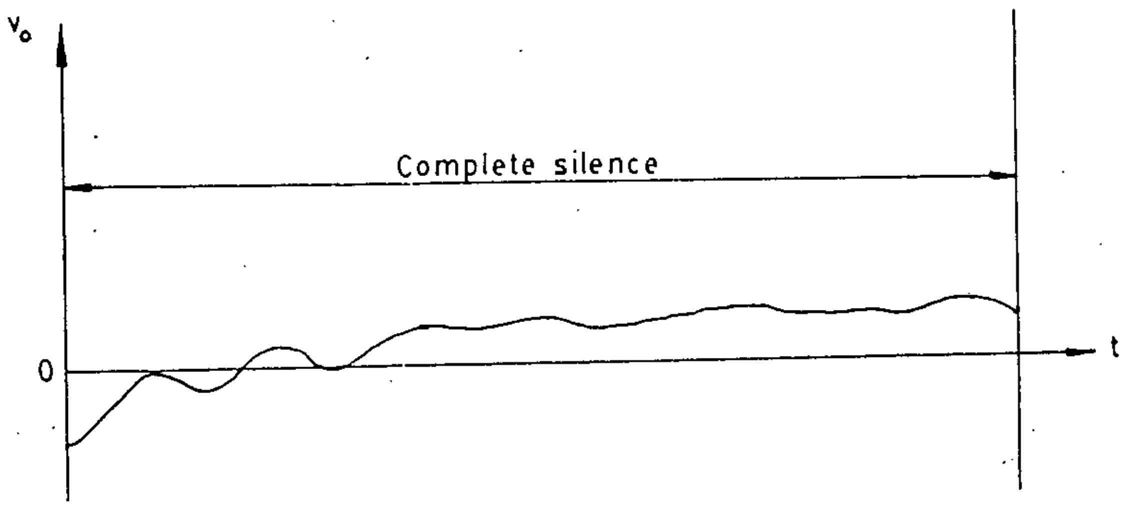
Figure 3.2 Block diagram of the experimental set-up.

The output of the tape recorder is played through a bandpass filter whose passband approximately range from 300 Hz to 3400 Hz. It is a common practice to restrict the speech band to a range of 300 Hz to 3400 Hz to make efficient use of the transmission channels.

Observation of the recorded speech signal on an oscilloscope shows that it is basically alternating in nature [Figure 3.3 (a)]. Determination of the presence of speech signal therefore, requires two threshold levels; of which one is positive and the other is negative. To avoid these two threshold levels for the single on-state (presence of speech), we have used full-wave rectified output of the alternating speech signal.



(a) A speech segment.



(b) A complete silence.

Figure 3.3 Electrical signals of conversational Bangla speech.

Crystal diodes are used in designing the full-wave rectifier. The forward voltage of the diodes usually range from 0.2 to 0.7 volt. But the amplitude of the recorded speech signal over telephone channels is quite low and not enough to drive the crystal diodes. Therefore, proper amplification of speech signal is required prior to full-wave rectification. A common audio amplifier and full-wave rectifier section is designed to serve the purpose. Figure 3.4 shows such a common audio-amplifier and full-wave rectifier section.

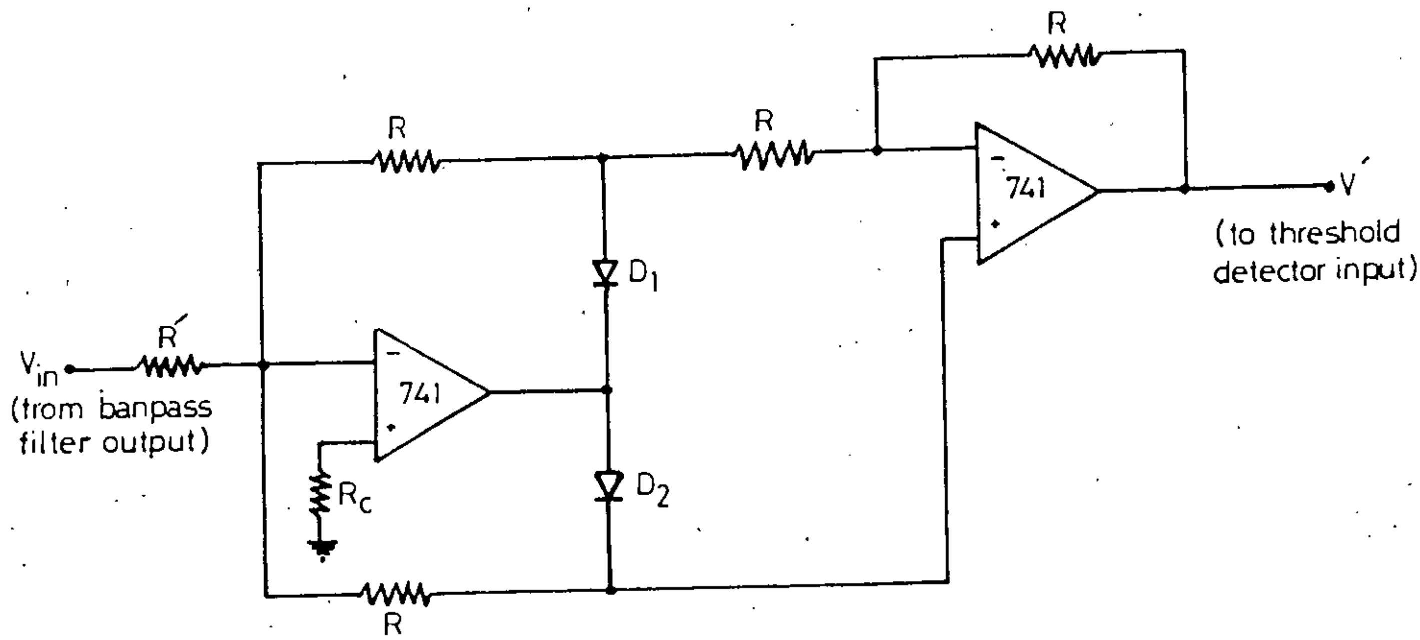


Figure 3.4 A common audio-amplifier and full-wave rectifier section.

The amplification factor is given by the relation

$$A = \frac{V'}{V_{in}} = \frac{R}{R'} \qquad \ldots \qquad \ldots (3.1)$$

The required amplification is achieved by selecting the proper values of resistors R and R´. For proper detection of the presence of speech signal an amplification factor of 10 is used.

3.4.3 Detection of presence or absence of speech

Since Time Assignment speech interpolation (TASI) systems were first introduced, several different speech detection algorithms have been developed. Most of these algorithms detect speech by comparing the signal levels, the energy of the signal or its envelope, the zero-crossing rate or combination of these, with preset threshold values [2]. In our detector, speech descrimination is based on the magnitude of the rectified signal voltage.

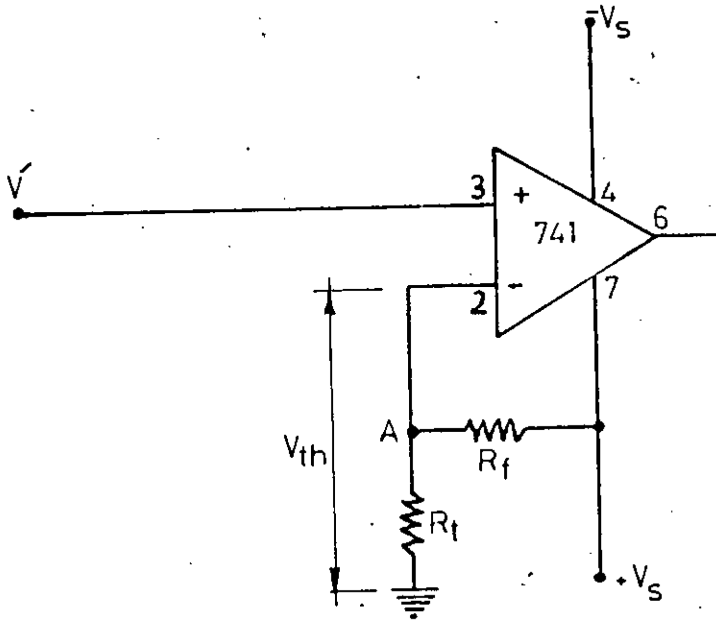
The speech-detector threshold should ideally be chosen low enough to pick up almost all the speech signal and high enough to avoid noise operation.

To set the threshold level, rectified and amplified speech signal is observed on an oscilloscope. By monitoring the audio speech the amplitude level of the noise signal corresponding to the silent portion of speech is observed carefully on the oscilloscope [Figure 3.3 (a) and (b)]. The threshold level is set slightly higher than the maximum noise level thus observed. The threshold detector stage is shown in figure 3.5.

The threshold voltage V_{th} (positive) may be set at the desired level by properly choosing the resistors R_t and R_f using the following relationship:

$$V_{th} = \frac{R_t}{R_t + R_f} \cdot V_s \qquad \ldots \qquad \ldots \qquad \ldots \qquad (3.2)$$

If the rectified signal level exceeds this set value, the output of the detector is set to $+V_s$ and if it falls below that, the output is set to $-V_s$. This polarity may be reversed if point A is connected to pin-3 and input is fed to pin-2 of the IC-741 in figure 3.5.



V'	V_o
$> V_{th}$	$+V_s$
$< V_{th}$	$-V_s$

<u>Figure</u> <u>3.5</u> Threshold detector stage.

3.5 Technique of data collection for speech analysis

The main purpose behind designing the speech detector is to be able to measure the short durations of the presence or absence of speech over telephone channel. Upto this point, we are able to convert the state of presence of speech to a voltage level $+V_s$ and the absence to another voltage level $-V_s$. As long as the speech is present or absent in the circuit, the output voltage $+V_s$ or $-V_s$ of the detector remains constant throughout that duration. For measuring these on-off durations we may plot the output from the detector with the help of a pen-recorder and then compute the time from the plot. But the main disadvantage of a pen-recorder is that the mechanical movements of the pen cannot coup with the fast changing states at the detector output. Moreover, a huge quantity of paper is required and computations become tedious. This is why, computer is used to measure these short durations of the on-off time.

3.5.1 Interface circuits

The output of the detector cannot be directly fed to the input port of the personal computer. To make the output compatible with PC an interface is placed in between the detector and the PC. For an IBM PC its input port specifications require -3 to -15 volts for binary 1 and +3 to +15 volts for binary 0. As the detector is designed by using an op-amp the detector output $+V_s$ and $-V_s$ may be set to such a value as to satisfy the above specifications. Thus, the detector output is directly fed to the personal computer without any additional interfacing device.

3.5.2 The computer technique for on-off data collection

Our speech detector is designed by using an operational amplifier and the value of V_s is set at +12 volts. The recorded speech is played through the speech detector unit to generate the on-off patterns. The output from the detector is fed to pin-6 of the RS-232 C port and pin-7 is connected to signal ground. The detector output is now digital in nature, having two states: one is +12 volts (binary 0), depending on the presence of speech and the other is -12 volts (binary 1), depending on the absence. In the true sense a computer cannot read any continuous signal continuously. As reading a data is a serial process, some time, though very little, is elapsed between two consecutive readings. Therefore, some sort of sampling takes place while reading a continuous signal. For the purpose of collecting on-off speech data, a computer program is prepared in PASCAL to read data from the port at an interval of

5 ms and to store the status of pin-6 in the buffer. Pin-6 senses the active low input. When speech is present the detector output becomes +12 volts. Therefore, pin-6 is active sensing it as low and as instructed in the program, it stores a '1' in the buffer to indicate the presence of speech. Otherwise, it stores a '0' in the buffer to indicate the absence. Thus a series of 1s are obtained for the presence of speech and a series of 0s for the absence. The presence and absence of speech are followed by one another in course of time. Later on counting the number of 1s in a series and multiplying it by the sampling time of 5 ms gives the duration of on-time for that particular time-segment. Similarly, counting the number of 0s in a series and multiplying it by the sampling time of 5 ms gives the duration of off-time for that particular time-segment. If the sampling time could be made as short as possible the collected data would represent almost continuous on-off patterns. But there are some limitations to choose it to such a value.

First of all, it is obvious that data can be read and then directly stored in a floppy diskette. But reading a data from the port and writing it directly in the diskette takes much longer time than writing it in the buffer. Thus sampling interval becomes greater than the desired value and the collected data loses accuracy. This is why, data are read from the port and are directly stored in the buffer. Once the buffer is full the stored data is transferred to the floppy diskette.

On the otherhand, if sampling interval is made very short, a higher memory capacity is required. But the computer buffer allows a maximum memory size of 32767 bytes for a PASCAL program. This is why, if shorter sampling interval is used, the buffer becomes full within a very short time and we are able to collect the data for a maximum duration of only 1 minute or less which is much lower than an average telephone call duration of about 3 minutes.

Considering all these factors, a 5 ms sampling interval is chosen. It is so chosen because it serves our purpose with a negligible error and it can store the data of about 2 minutes and 38 seconds (i.e. 32767 bytes X 5 ms), which is very close to an average telephone call duration of about 3 minutes. Moreover, the same 5 ms sampling interval has been used in the study of on-off patterns of English speech [2]. Therefore, if we use a sampling interval of 5 ms, we can compare the results of Bangla speech with those of English one on the same basis.

The flow diagram of the speech detection algorithm is given in figure 3.6. Accordingly, a computer program named 'SAIF.PAS' is prepared in PASCAL to read data from the RS-232 C serial port of a PC. A printout of the program is given in Appendix C.

Thus the speech detector generates the on-off patterns and the designed computer program collects the data successfully. A portion of the collected data is shown in table 3.1.

The next chapter deals with calculation of different speech temporal parameters of conversational Bangla speech and their statistical analysis.

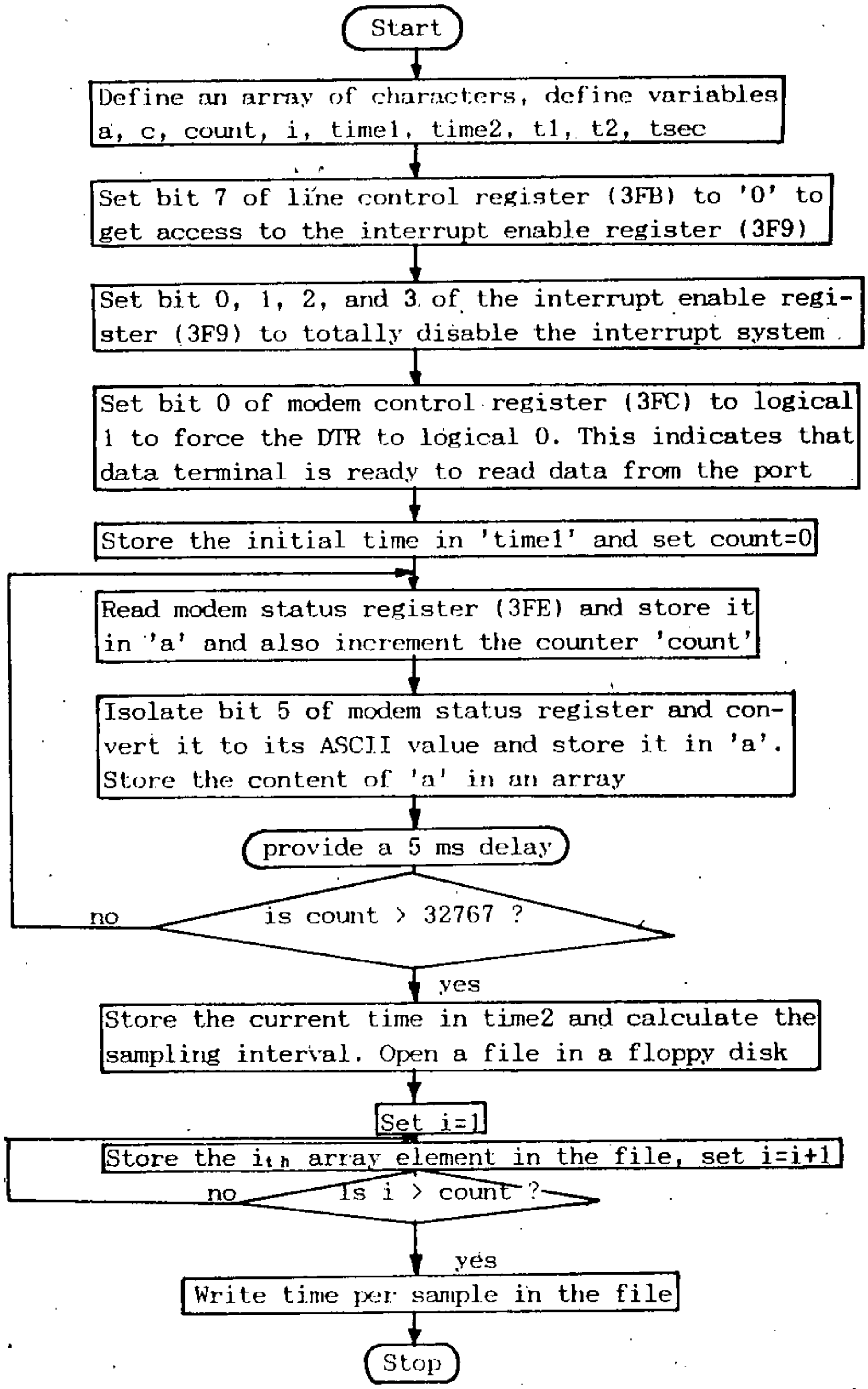


Figure 3.6 Flow diagram of the speech detection algorithm.

Table 3.1 On-off speech data represented by series of 1s and 0s respectively

```
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000010110111010100000100111111111111110101100111 0
01110110100111000100110001011000000000000000000000000011101111000000101100000
00000000010010000000100010000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
000000000110111011110111111111000010001000000001001100000110101010101111111111
11000000100010000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000100101000001101101110001100100000000000000000000011
11000101101111110100000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000100000000000000000000000000000000000000000000000
00000000000011000001000101111111111011101110000000000000000000000000000000000000
00000000101111110111111011101010000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000001001011111111111100000000000000000000000000000000001
```

Chapter 4

Bangla speech data processing and analysis

4.1 Introduction

The process of collecting the on-off speech data is already described in the previous chapter and we have now been able to collect the on-off speech data for conversational Bangla speech from the output of the speech detector.

This chapter describes the on-off data processing technique of conversational Bangla speech and the analysis of the data. The effect of fill-in and hangover on different speech temporal parameters are also studied.

4.2 Effect of threshold level on speech patterns

The speech patterns from a speech detector are strongly influenced by choice of threshold.

To see how low a threshold is required for the first criterion, the effect of threshold variation on the continuous speech spurt and pause distribution is studied. This distribution is a cumulative distribution function which shows the percent of total number of spurts or pauses (less than or equal to a particular time duration) along the ordinate and the duration of spurts or pauses along the abscissa.

First of all, a higher threshold level is set (2.164 volts) and then it is lowered step by step. The idea is, if a point is reached where the data remain substantially unchanged as the threshold is lowered, then that threshold would be considered sufficiently low to cause operation on most of the speech.

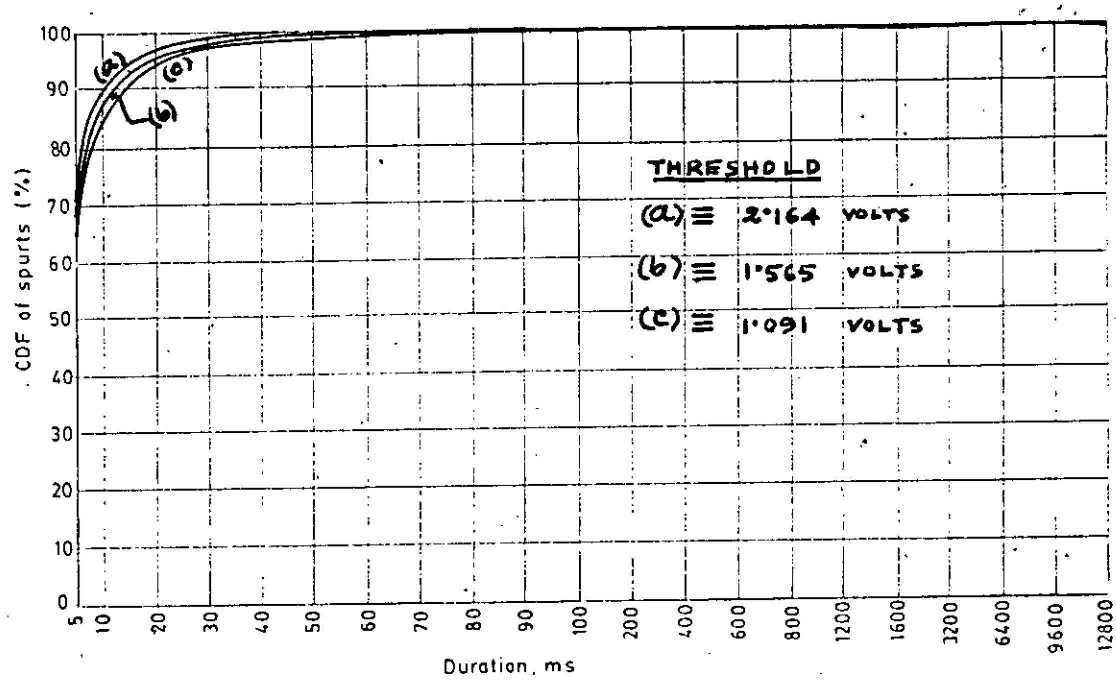
The three samples (correspoding to three thresholds) from a particular speaker (whose voice is low enough) are analyzed to see the effect of threshold variation on spurt and pause distributions [Figure 4.1].

The spurt distributions change appreciably as the threshold is varied, showing shorter spurts with higher thresholds.

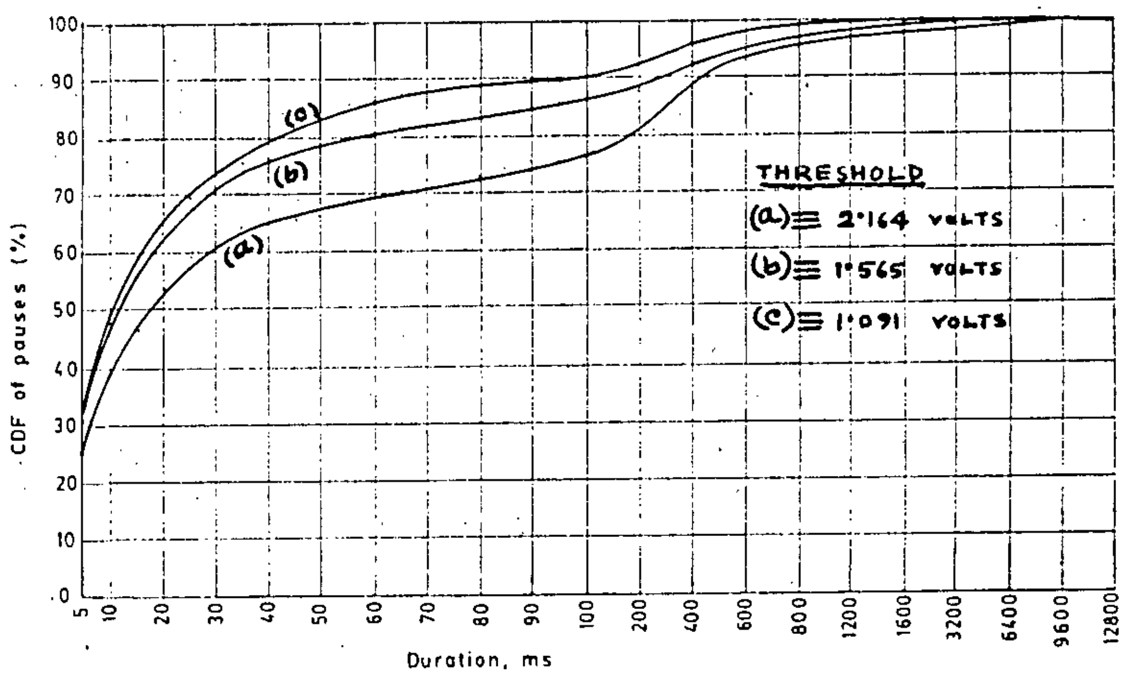
The pause distribution tells only the part of the story. The total number of pauses almost triples as the threshold is lowered from (a) to (c) going from 913 to 2569 pauses in a continuous speech sample of about 2 minutes and 38 seconds.

If the threshold is further lowered, the spurt and pause distributions do not stabilize, and neither do the percent time talking (i.e. speech activity). It then appears that the threshold should be still lowered in order to pick up most of the speech. However, it must be set slightly higher than the noise level. Otherwise, the result would be erroneous, because a silent interval will then appear to be a talkspurt. To estimate a lower bound, the original tapes are played and the detector output is compared with the impressions of the sound on the tape. This process has led us to choose a threshold level of 1.091 volts (out of 1.091, 1.565 and 2.164 volts). It should be noted that the original speech signal is first amplified by a voltage amplification factor of 10 and the threshold is selected at the amplified output.

The threshold is thus an unavoidable compromise between too much noise and too much lost speech.



(a) Effect of threshold on spurt distributions.



(b) Effect of threshold on pause distributions.

<u>Figure 4.1</u> Effect of threshold variation on spurt and pause distributions.

4.3 Data processing

The collected data from the output of the speech detector do not represent the true on-off patterns as the intersyllabic gaps are still present in them. To obtain the true on-off patterns some modifications are to be made on the data already collected.

Usually the first step in data processing is to throw out all spurts which are less than or equal to a throwaway time. In our analysis we have not considered this time on the assumption that our speech data contain less noise because of not recording it from the original telephone circuit. As we do not intend to perform a throwaway operation the first step of our data processing is to fill-in all the gaps which are less than or equal to a fill-in time. These gaps are considered as part of continuous speech segments. The speech data after fill-in operation represent the true on-off patterns.

If hangover operation is performed on the original data, the on-off patterns corresponding to a conventional detector with that particular hangover time is obtained.

4.4 Selection of fill-in and hangover time for conversational Bangla speech

The purpose of fill-in or hangover is to discard the intersyllabic gaps within the same speech segment. The fill-in or hangover time should be chosen just long enough to bridge the longest intersyllabic gap within a continuous speech segment. Therefore, for the analysis of conversational Bangla speech, we have to measure the longest intersyllabic gap. For this purpose data are collected from different speech samples at the fastest possible sampling rate (0.17 ms). The on-off patterns are observed on the computer screen. After a careful consideration of all the gaps within the speech segments, the intersyllabic gaps appear to lie between 2 and 11.4 ms. The original data have been collected using a 5 ms sampling interval. Therefore, we have two choices to select a fill-in time; either of 10 ms or of 15 ms. We choose the fill-in time to be 10 ms as it is close to the maximum intersyllabic gap of 11.4 ms. As already stated in Chapter 3, it is logical to choose the hangover time equal to the maximum intersyllabic gap. Therefore, the hangover time is also chosen to be 10 ms.

The modified form of table 3.1 (chapter 3) after applying a 10 ms fill-in is shown in table 4.1 and that after applying a 10 ms hangover is shown in in table 4.2.

4.5 Bangla speech temporal parameters under different conditions

The on-off data are collected by the computer in the form of series of 1s and 0s. The series of 1s represents the presence of speech and that of 0s represents the absence as shown in table 3.1 in chapter 3. Now counting the number of 1s in a series and multiplying it by the sampling interval of 5 ms gives the duration of on-time for that particular time-segment. Similarly, counting the number of 0s in a series and following the same procedure gives the duration of off-time for that particular time-segment. Thus the on-off patterns in terms of 1s and 0s are converted to on-off patterns in terms of the durations of time. With these on-off patterns different speech temporal parameters such as speech activity, average spurt time, average pause time and talkspurt rate of conversational Bangla speech are calculated for male and female voices separately with no modifications of the data. The same parameters are calculated with a 10 ms fill-in and also with a 10 ms hangover. The cumulative distribution functions (CDF) of talkspurts and pauses are also shown graphically. The CDF shows the percent of total number of spurts or pauses (less than or equal to a particular time duration) along the ordinate

Table 4.1 Modified data of table 3.1 with a 10 ms fill-in

0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000111111111110000011111111111111111111111111
1111111111111000111110001110000000000000000111111100000011100000000
0000000000111000000010001000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000001111111111111111110000100010000000111110000111111111111111
1100000001000100000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000111110000011111111000111100000000000000011
1100011111111111100000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000001000000000000000000000000000000000000000
0000000000001100000100011111111111111111110000000000000000000000000
0000000011111111111111111110000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000111111111111110000000000000000000000000000001

Table 4.2 Modified data of table 3.1 with a 10 ms hangover

0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000111111111111110001111111111111111111111111
1111111111111111011111110111111000000000000000011111111110000111111000
0000000000111111000011101110000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000001111111111111111111111001110111000001111111001111111111111111
1111000001110111000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000011111111000111111111111101111111000000000011
1111011111111111111100000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000011100000000000000000000000000000000000000000
0000000000001111000111011111111111111111111100000000000000000000000000
0000000011111111111111111111110000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000001111111111111111100000000000000000000000000001

and the duration of spurts or pauses along the abscissa.

The following defining equations are used to calculate the different speech temporal parameters:

$$SA = \frac{\sum_{i=1}^{m} S_i}{\sum_{i=1}^{m} S_i + \sum_{j=1}^{n} P_j} \times 100\ \% \qquad \ldots \quad \ldots \quad (4.1)$$

where,

SA= speech activity in %

m= total number of spurts

S_i = duration of the i_{th} spurt

n= total number of pauses and

P_j = duration of the j_{th} pause

$$AST = \frac{\sum_{i=1}^{m} S_i}{m} \qquad \ldots \quad \ldots \quad (4.2)$$

where,

AST= average spurt time

m= total number of spurts and

S_i = duration of the i_{th} spurt

$$APT = \frac{\sum_{j=1}^{n} P_j}{n} \qquad \ldots \quad \ldots \quad (4.3)$$

where,

APT= average pause time

n= total number of pauses and

P_j = duration of the j_{th} pause

$$TR = \frac{m}{\sum_{i=1}^{m} S_i + \sum_{j=1}^{n} P_j} \qquad \ldots \quad \ldots \quad (4.4)$$

where,

TR= talkspurt rate per unit time

m= total number of spurts

S_i = duration of the i_{th} spurt

n= total number of pauses and

P_j = duration of the j_{th} pause

4.5.1 Speech temporal parameters without fill-in or hangover

Speech temporal parameters of conversational Bangla speech are calculated from the original data without any modification. The parameter values do not represent the true valuse of conversational Bangla speech. They are calculated just to see the extent of deviation of the speech parameters from the true ones. Table 4.1 lists the speech parameters for male voice and table 4.2 for female. Tables 4.3 and 4.4 show the maximum, minimum and average values of the speech temporal parameters for male and female voices respectively. The average values are calculated using the data of a group of 10 males or 10 females using the equations (4.1), (4.2), (4.3) and (4.4). The overall speech parameter values (male and female combined) are also calculated using the aforementioned procedure and are listed in table 4.5. The cumulative distribution functions of spurts for male and female voices are shown in figures 4.2 and 4.3 respectively. The same of pauses are shown in figures 4.4 and 4.5. It should be noted here that the lower and upper bounds are the curves for individual samples and the average curves cosider the data of 10 subjects as a whole in a group. Figures 4.6 and 4.7 show the overall (male and female combined) CDFs of spurts and pauses respectively.

Subjects	Speech activity (%)	Average spurt time (ms)	Average pause time (ms)	Talkspurt rate (per min.)
M-1	08.63	05.92	62.69	0874.6
M-2	26.59	09.13	25.19	1748.4
M-3	25.46	08.50	24.90	1796.0
M-4	19.16	13.21	55.74	0870.2
M-5	16.88	16.41	80.87	0617.1
M-6	14.92	12.31	70.19	0727.3
M-7	18.75	22.01	95.49	0510.9
M-8	10.78	11.92	98.63	0542.8
M-9	08.24	11.25	125.27	0439.5
M-10	16.90	13.13	64.55	0772.4

Table 4.1 Speech temporal parameters for male voice without fill-in or hangover

Subjects	Speech activity (%)	Average spurt time (ms)	Average pause time (ms)	Talkspurt rate (per min.)
F-1	17.11	09.87	47.85	1039.4
F-2	25.90	06.73	19.27	2307.3
F-3	24.41	08.28	25.65	1768.2
F-4	21.34	10.96	40.43	1167.9
F-5	14.01	08.93	54.84	0941.2
F-6	11.73	06.77	50.96	1039.4
F-7	16.20	08.23	42.56	1181.5
F-8	09.63	07.97	74.73	0725.5
F-9	09.08	07.84	78.48	0695.1
F-10	12.35	09.68	68.70	0765.4

Table 4.2 Speech temporal parameters for female voice without fill-in or hangover

Speech parameters	Maximum	Minimum	Average of 10 speakers
Speech activity	26.59%	08.63%	16.63%
Average spurt time	22.01 ms	05.92 ms	11.21 ms
Average pause time	125.27 ms	24.90 ms	56.21 ms
Talkspurt rate	1796.0 per min.	439.5 per min.	889.92 per min.

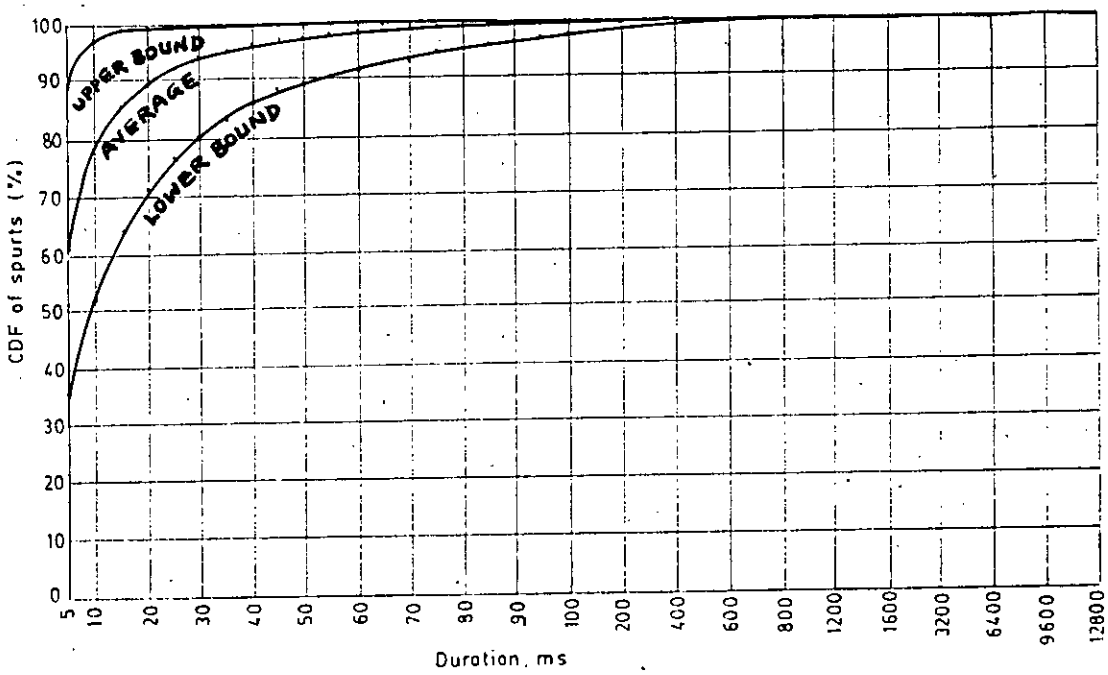
Table 4.3 **Maximum and minimum limits for speech temporal parameters for male voice without fill-in or hangover**

Speech parameters	Maximum	Minimum	Average of 10 speakers
Speech activity	25.90%	09.08%	16.18%
Average spurt time	10.96 ms	06.73 ms	08.34 ms
Average pause time	78.48 ms	19.27 ms	43.25 ms
Talkspurt rate	2307.3 per min.	695.1 per min.	1163.09 per min.

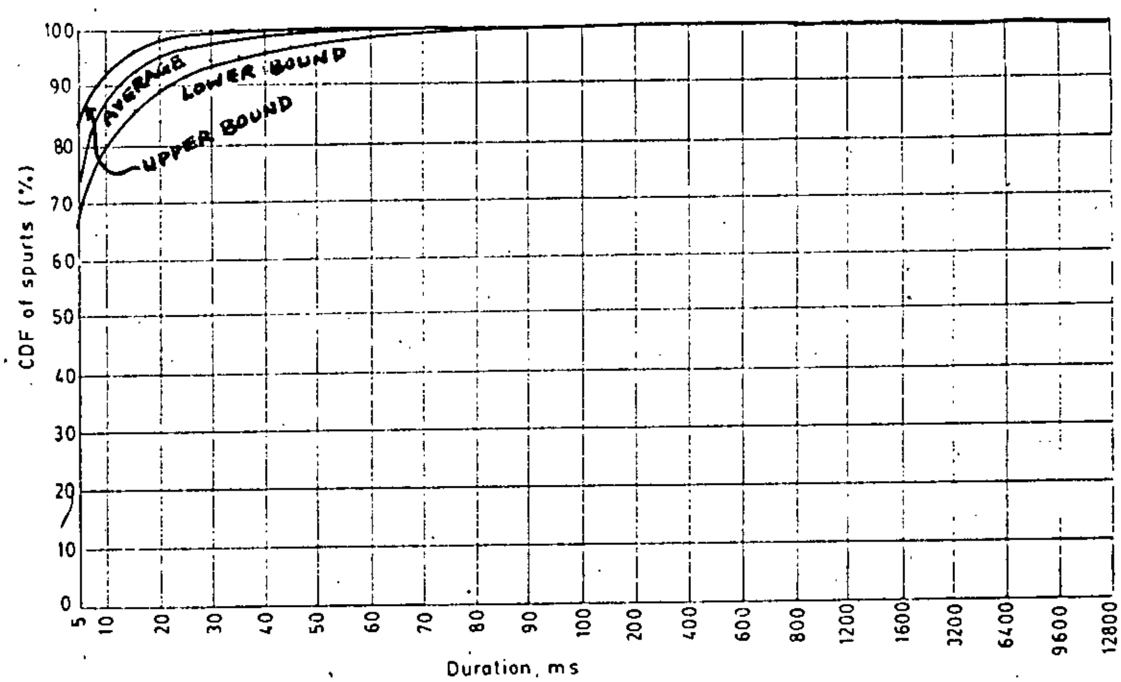
Table 4.4 **Maximum and minimum limits for speech temporal parameters for female voice without fill-in or hangover**

Speech activity	-	16.40%
Average spurt time	-	09.59 ms
Average pause time	-	48.86 ms
Talkspurt rate	-	1026.50 per min.

Table 4.5 Overall on-off speech statistics (male and female combined) without fill-in or hangover



<u>Figure</u> <u>4.2</u> CDF of spurts for male voice without fill-in or hangover.



<u>Figure</u> <u>4.3</u> CDF of spurts for female voice without fill-in or hangover.

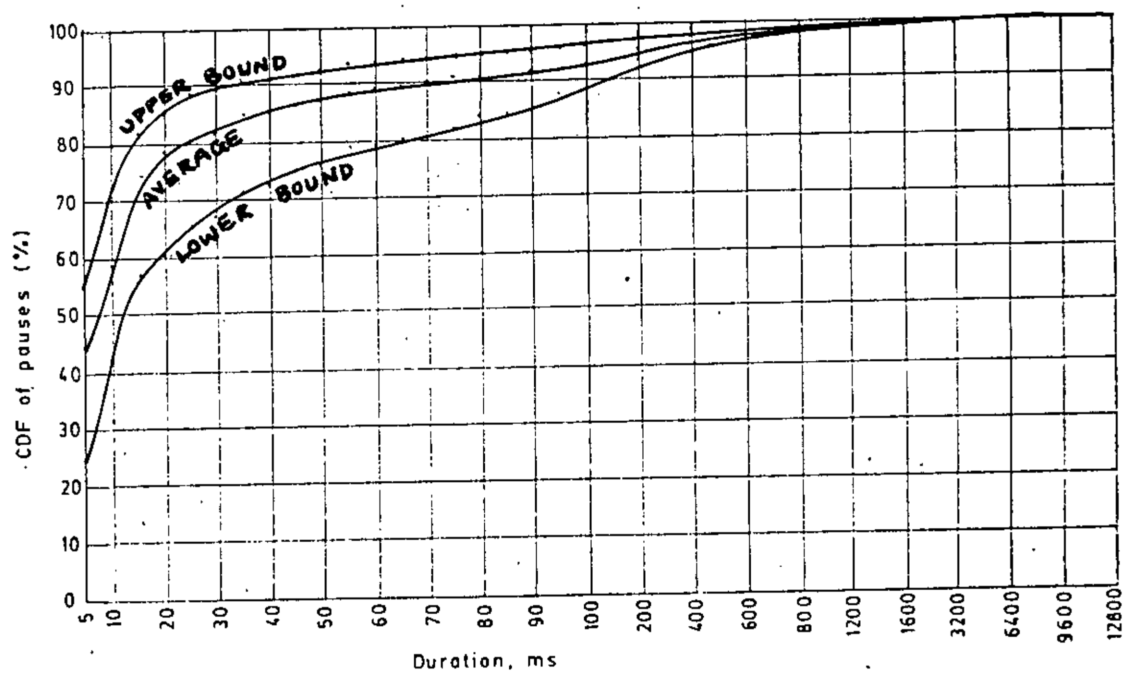


Figure 4.4 CDF of pauses for male voice without fill-in or hangover.

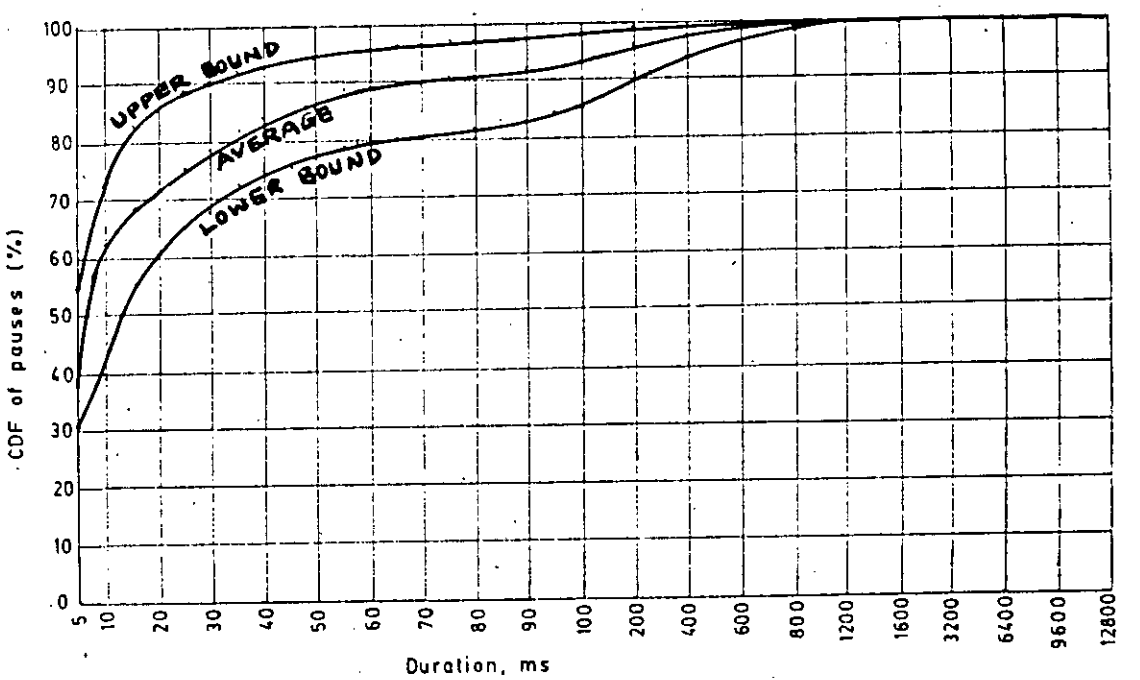


Figure 4.5 CDF of pauses for female voice without fill-in or hangover.

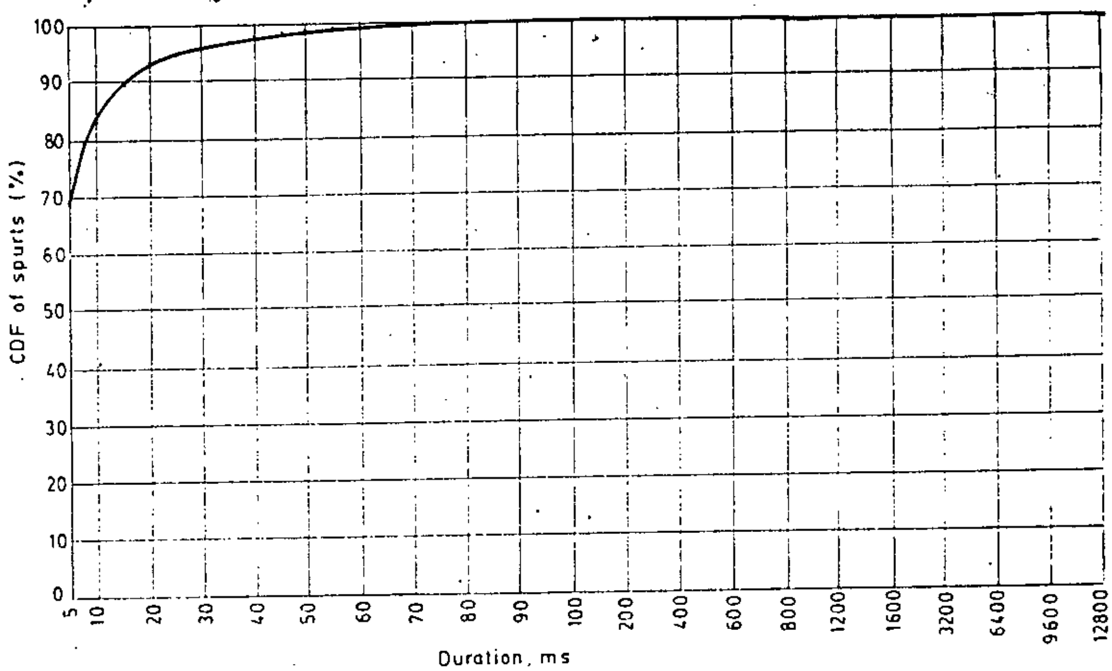


Figure 4.6 Overall CDF of spurts without fill-in or hangover.

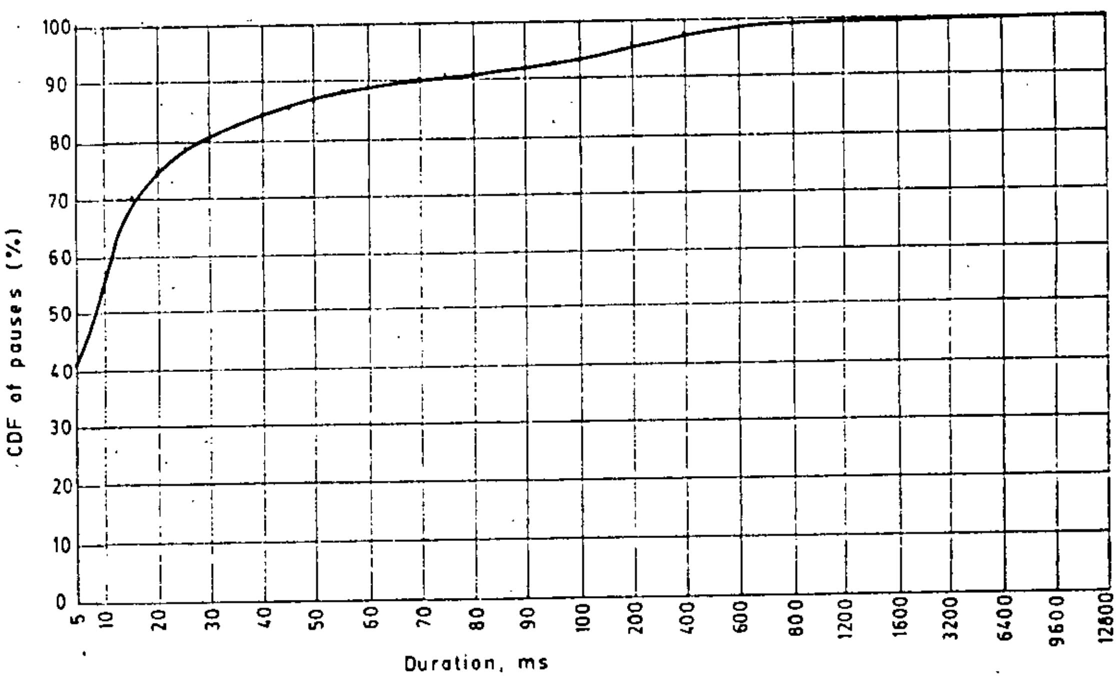


Figure 4.7 Overall CDF of pauses without fill-in or hangover.

4.5.2 Speech temporal parameters with a 10 ms fill-in

Speech temporal parameters are calculated using the modified form of the original data after applying a 10 ms fill-in time. The calculated speech temporal parameters represent the true on-off characteristics of conversational Bangla speech. Different speech parameter values are shown in tables 4.6, 4.7, 4.8, 4.9. and 4.10. The CDFs of spurts and pauses are shown graphically in figures 4.8, 4.9, 4.10, 4.11, 4.12 and 4.13.

4.5.3 Speech temporal parameters with a 10 ms hangover

Speech temporal parameters are also calculated using the data modified by a 10 ms hangover time. The calculated characteristics correspond to those obtainable from a conventional detector having a hangover time of 10 ms. The speech parameter values are shown in tables 4.11, 4.12, 4.13, 4.14 and 4.15. The CDFs of spurts and pauses are shown graphically in figures 4.14, 4.15, 4.16, 4.17, 4.18 and 4.19.

Subjects	Speech activity (%)	Average spurt time (ms)	Average pause time (ms)	Talkspurt rate (per min.)
M-1	11.52	11.66	89.54	0592.9
M-2	37.29	35.77	60.15	0625.5
M-3	35.52	26.51	48.15	0803.9
M-4	25.10	46.84	139.76	0321.6
M-5	21.29	70.04	259.46	0182.4
M-6	20.09	48.84	194.23	0246.8
M-7	22.39	86.93	302.01	0154.6
M-8	14.16	35.91	217.70	0236.6
M-9	10.81	30.27	249.78	0214.2
M-10	22.03	43.06	152.44	0306.9

Table 4.6 Speech temporal parameters for male voice with a 10 ms fill-in

Subjects	Speech activity (%)	Average spurt time (ms)	Average pause time (ms)	Talkspurt rate (per min.)
F-1	22.39	24.93	86.44	538.7
F-2	39.61	27.58	42.06	861.7
F-3	33.34	21.72	43.43	921.1
F-4	27.15	26.90	72.20	605.7
F-5	19.11	23.89	101.16	480.1
F-6	13.43	21.54	138.92	373.9
F-7	22.61	23.48	80.40	577.5
F-8	15.85	14.87	78.96	639.8
F-9	11.95	17.14	126.43	418.2
F-10	16.51	26.16	132.28	378.7

<u>**Table 4.7**</u> <u>**Speech temporal parameters for female voice with a 10 ms fill-in**</u>

Speech parameters	Maximum	Minimum	Average of 10 speakers
Speech activity	37.29%	10.81%	22.02%
Average spurt time	86.93 ms	11.66 ms	35.85 ms
Average pause time	302.01 ms	48.15 ms	126.99 ms
Talkspurt rate	803.9 per min.	154.6 per min.	368.54 per min.

Table 4.8 Maximum and minimum limits for speech temporal parameters for male voice with a 10 ms fill-in

Speech parameters	Maximum	Minimum	Average of 10 speakers
Speech activity	39.61%	11.95%	22.19%
Average spurt time	27.58 ms	14.87 ms	22.98 ms
Average pause time	138.92 ms	42.06 ms	80.58 ms
Talkspurt rate	921.1 per min.	373.9 per min.	579.54 per min.

Table 4.9 Maximum and minimum limits for speech temporal parameters for female voice with a 10 ms fill-in

Speech activity	-	22.11%
Average spurt time	-	27.98 ms
Average pause time	-	98.63 ms
Talkspurt rate	-	474.04 per min.

Table 4.10 Overall on-off speech statistics (male and female combined) with a 10 ms fill-in

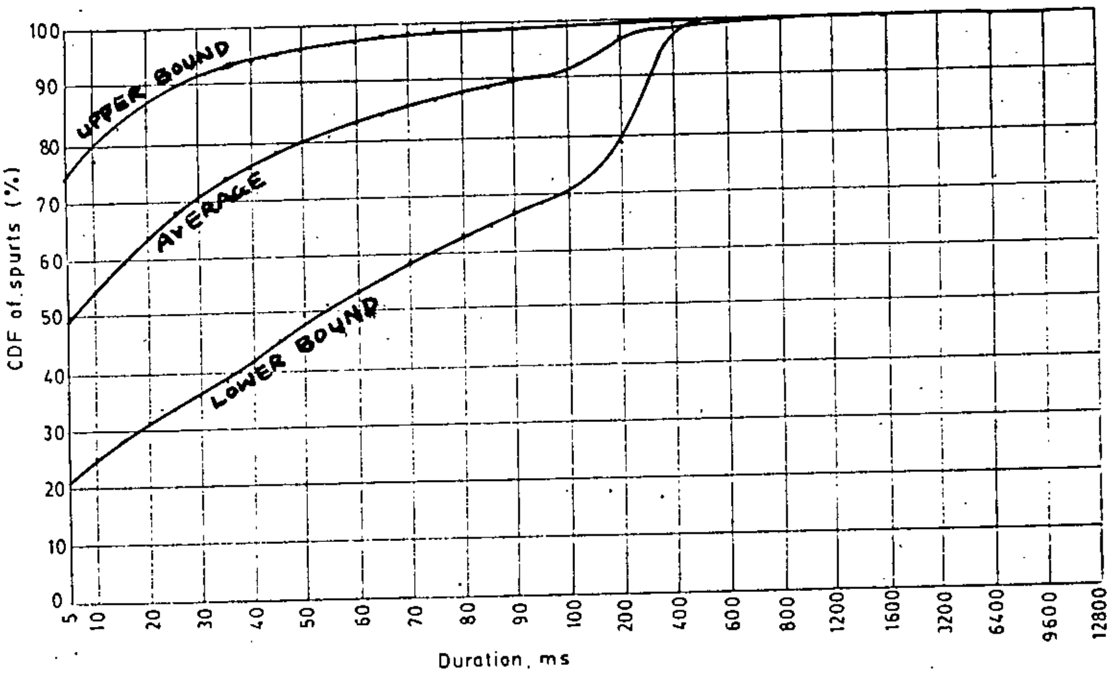


Figure 4.8 CDF of spurts for male voice with a 10 ms fill-in.

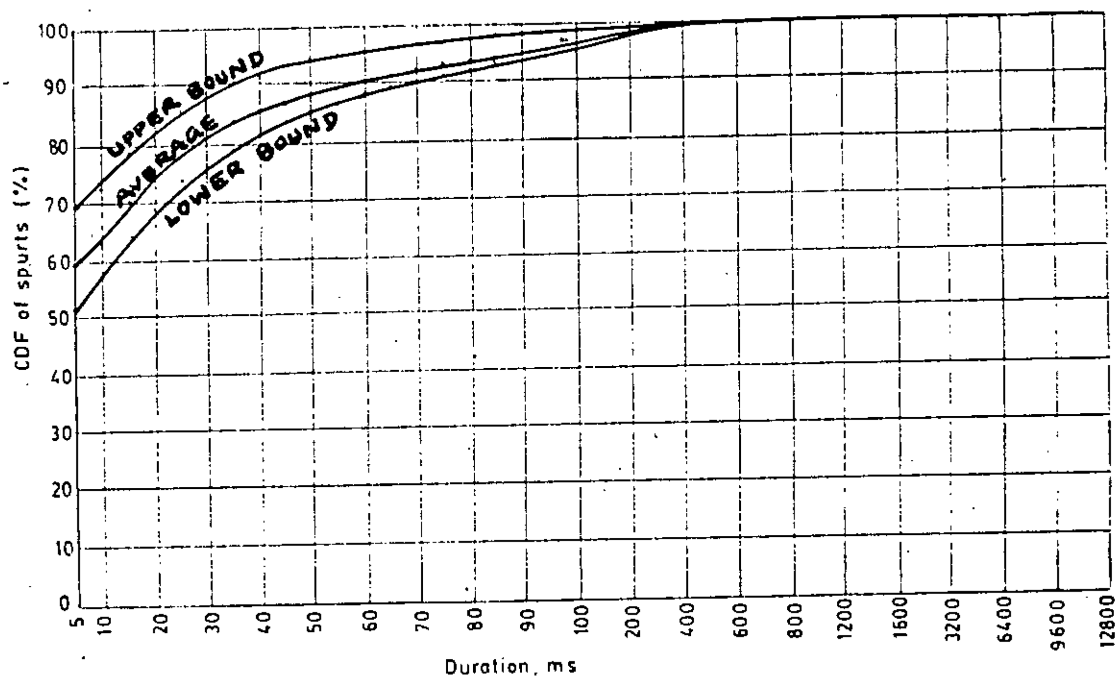


Figure 4.9 CDF of spurts for female voice with a 10 ms fill-in.

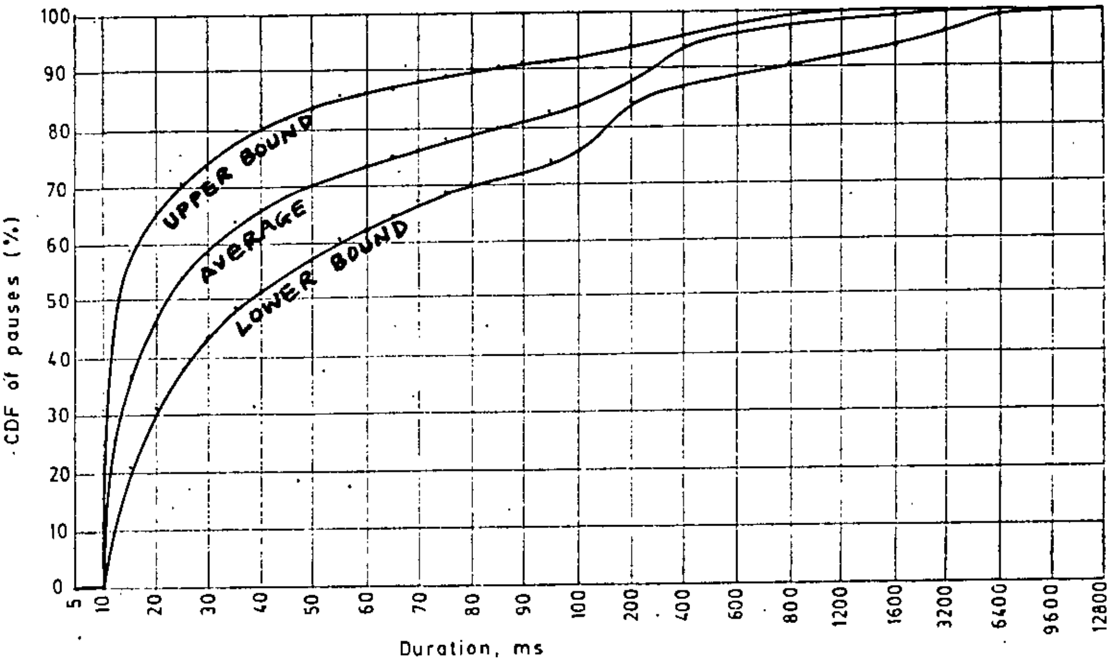


Figure 4.10 CDF of pauses for male voice with a 10 ms fill-in.

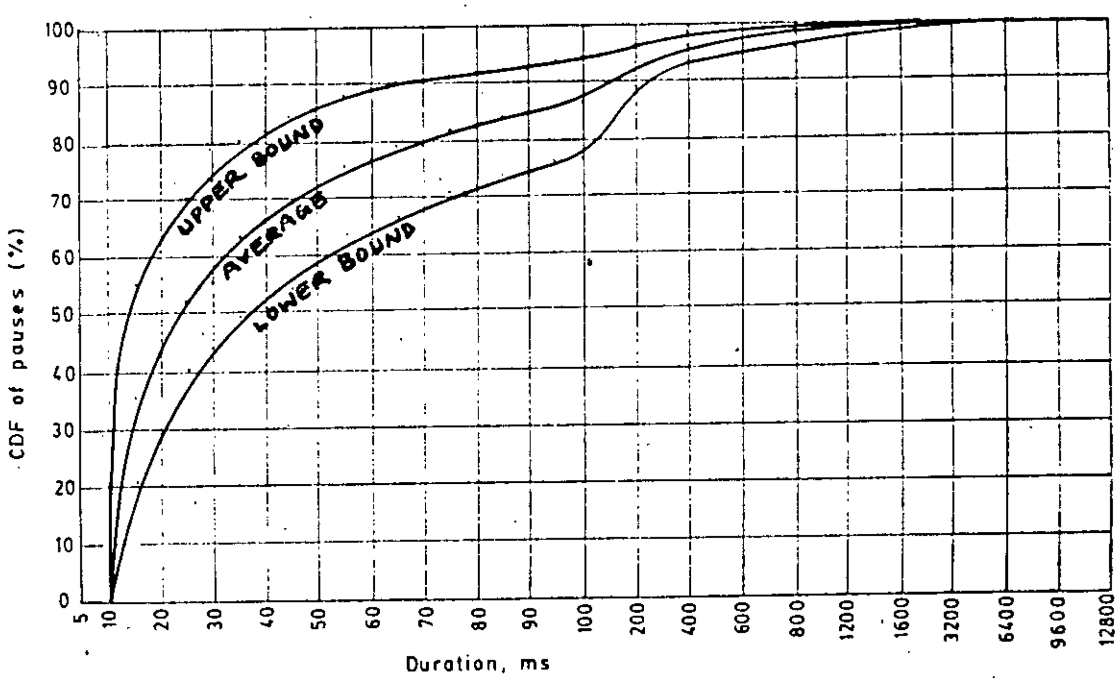


Figure 4.11 CDF of pauses for female voice with a 10 ms fill-in.

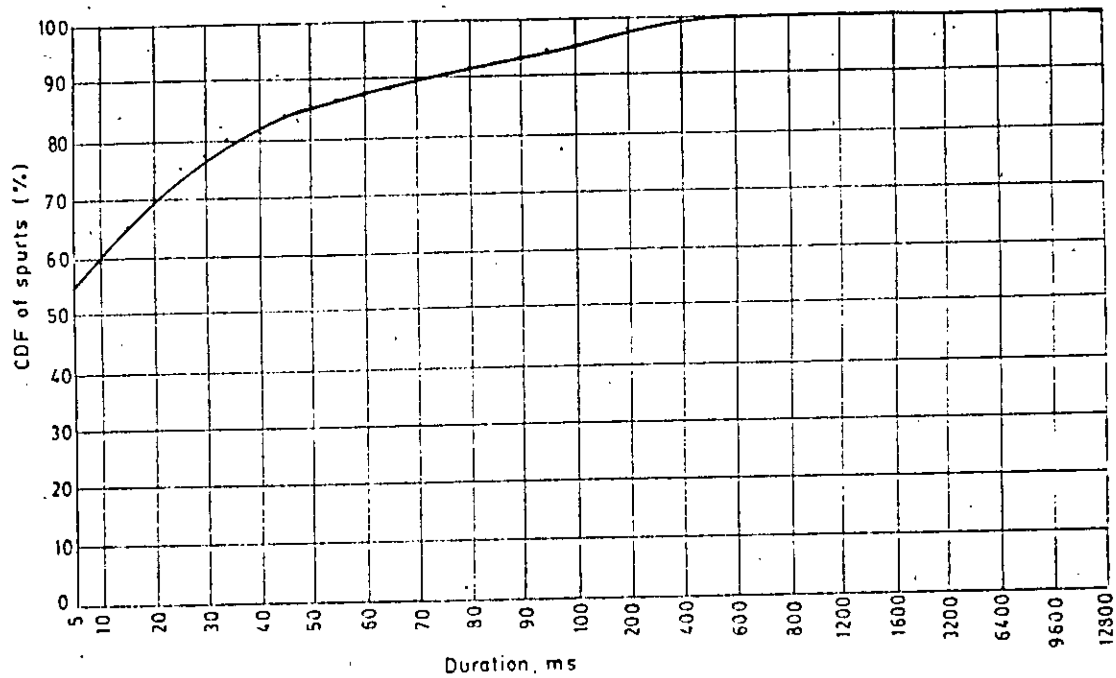


Figure 4.12 Overall CDF of spurts with a 10 ms fill-in.

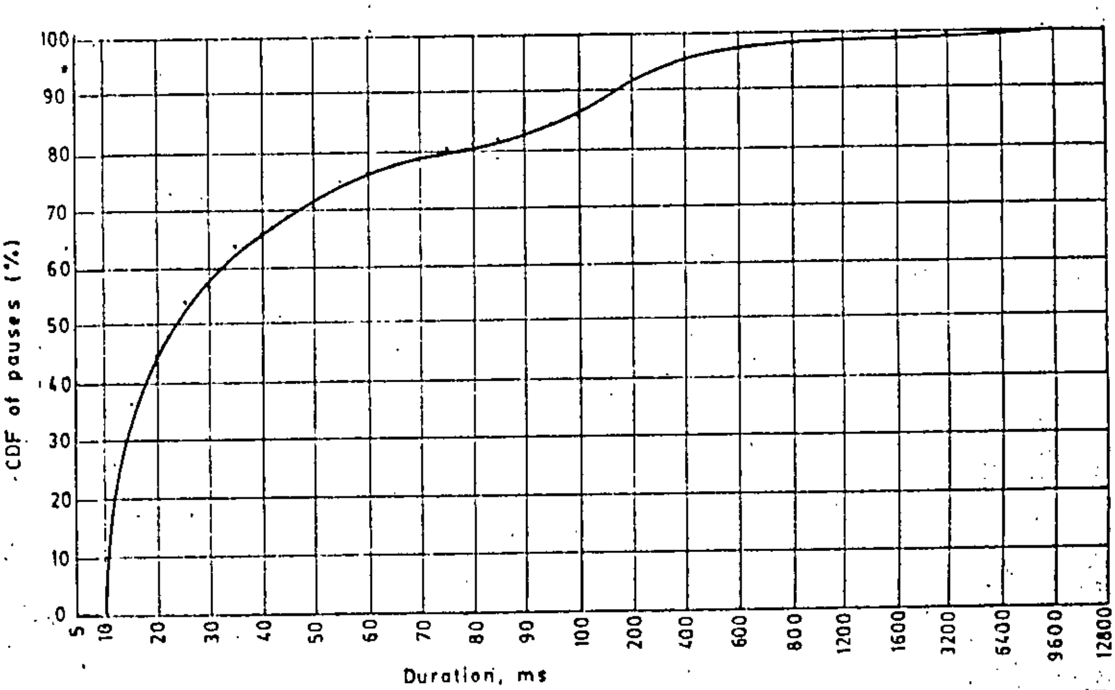


Figure 4.13 Overall CDF of pauses with a 10 ms fill-in.

Subjects	Speech activity (%)	Average spurt time (ms.)	Average pause time (ms)	Talkspurt rate (per min.)
M-1	21.40	21.66	79.54	592.9
M-2	47.72	45.77	50.15	625.5
M-3	48.91	36.51	38.15	803.9
M-4	30.46	56.84	129.76	321.6
M-5	24.32	80.02	249.46	182.4
M-6	24.21	58.84	184.23	246.8
M-7	24.96	96.91	292.01	154.6
M-8	18.10	45.91	207.70	236.6
M-9	14.38	40.27	239.78	214.2
M-10	27.14	53.06	142.44	306.9

Table 4.11 Speech temporal parameters for male voice with a 10 ms hangover

Subjects	Speech activity (%)	Average spurt time (ms)	Average pause time (ms)	Talkspurt rate (per min.)
F-1	31.36	34.93	76.44	538.7
F-2	53.97	37.58	32.06	861.7
F-3	48.69	31.72	33.43	921.1
F-4	37.24	36.89	62.20	605.7
F-5	27.11	33.90	91.23	479.8
F-6	19.66	31.54	128.92	373.9
F-7	32.23	33.48	70.40	577.5
F-8	26.51	24.86	68.96	639.8
F-9	18.91	27.13	116.43	418.2
F-10	22.82	36.16	122.28	378.7

Table 4.12 Speech temporal parameters for female voice with a 10 ms hangover

Speech parameters	Maximum	Minimum	Average of 10 speakers
Speech activity	47.72%	14.38%	28.16%
Average spurt time	96.91 ms	21.66 ms	45.85 ms
Average pause time	292.01 ms	38.15 ms	116.99 ms
Talkspurt rate	803.9 per min.	154.6 per min.	368.54 per min.

<u>Table 4.13</u> <u>Maximum and minimum limits for speech temporal parameters for male voice with a 10 ms hangover</u>

Speech parameters	Maximum	Minimum	Average of 10 speakers
Speech activity	53.97%	18.91%	31.85%
Average spurt time	37.58 ms	24.86 ms	32.98 ms
Average pause time	128.92 ms	32.06 ms	70.59 ms
Talkspurt rate	921.1 per min.	373.9 per min.	579.51 per min.

<u>Table 4.14</u> <u>Maximum and minimum limits for speech temporal parameters for female voice with a 10 ms hangover</u>

75029

Speech activity	-	30.00%
Average spurt time	-	37.98 ms
Average pause time	-	88.63 ms
Talkspurt rate	-	474.02 per min.

Table 4.15 Overall on-off speech statistics (male and female combined) with a 10 ms hangover

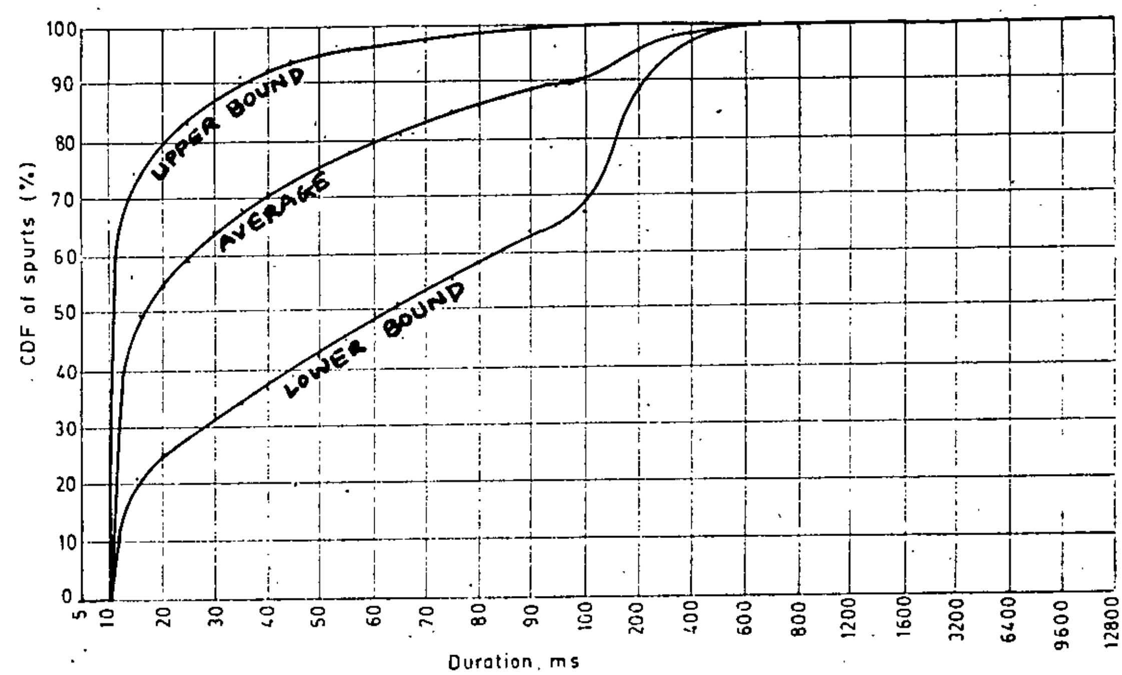


Figure 4.14 CDF of spurts for male voice with a 10 ms hangover.

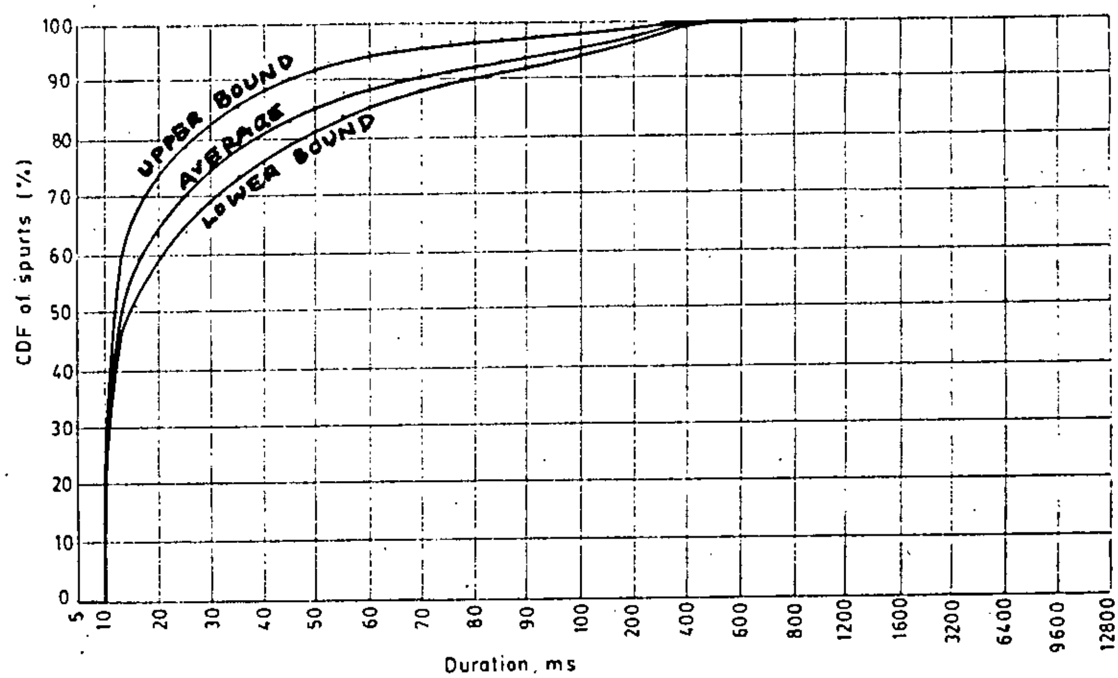


Figure 4.15 CDF of spurts for female voice with a 10 ms hangover.

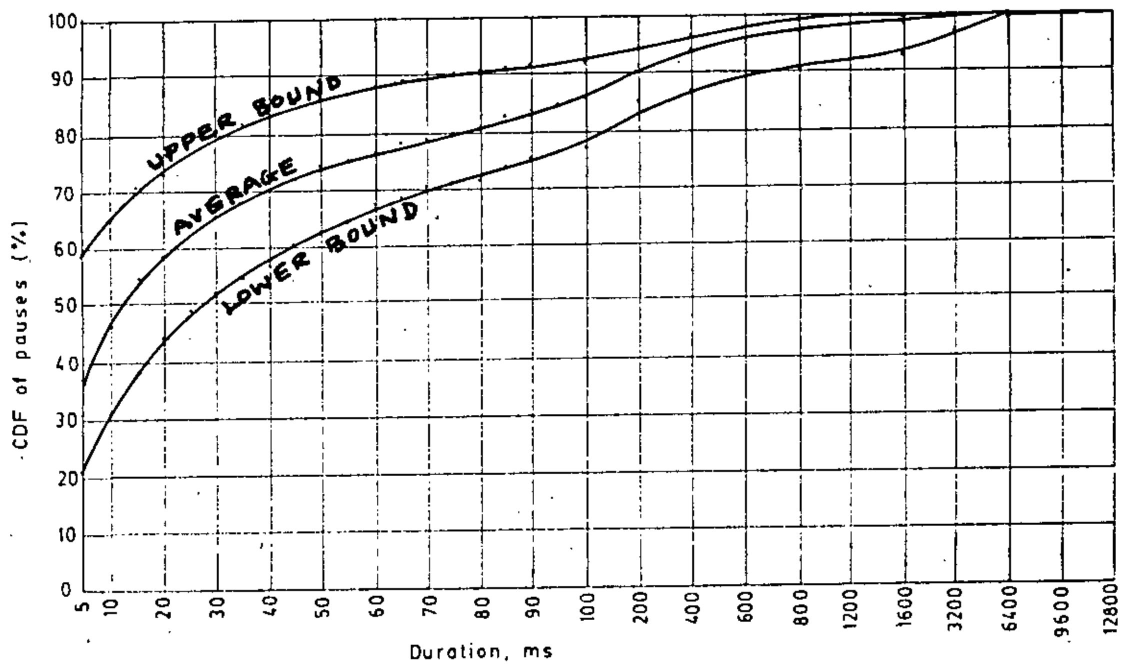


Figure 4.16 CDF of pauses for male voice with a 10 ms hangover.

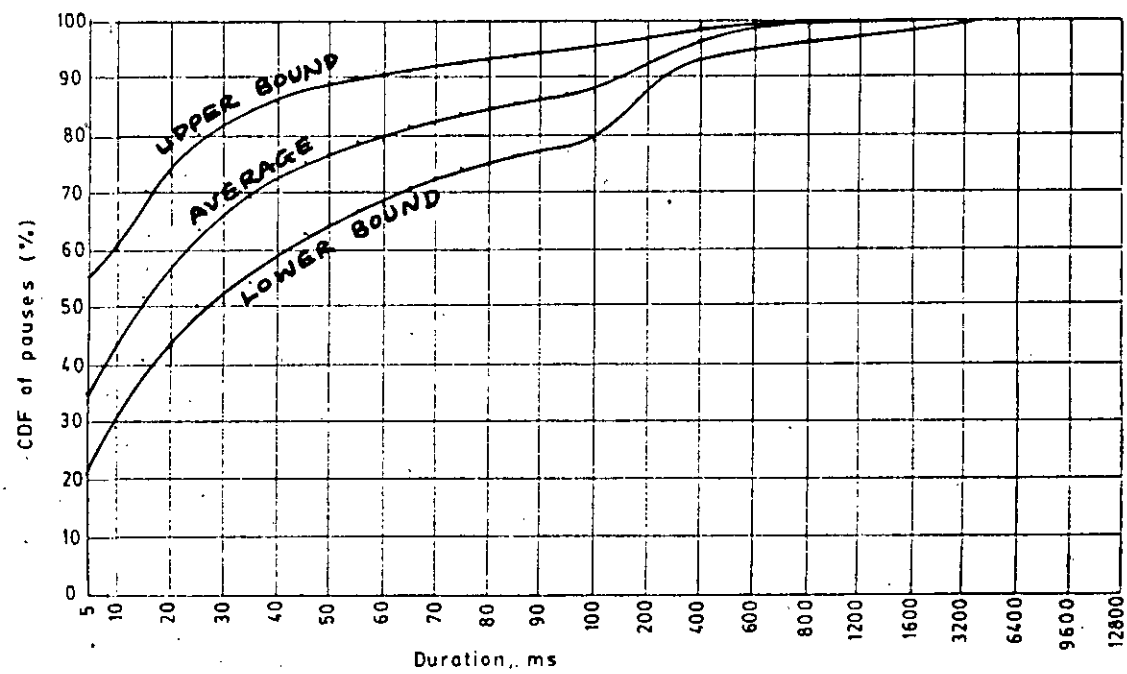


Figure 4.17 CDF of pauses for female voice with a 10 ms hangover.

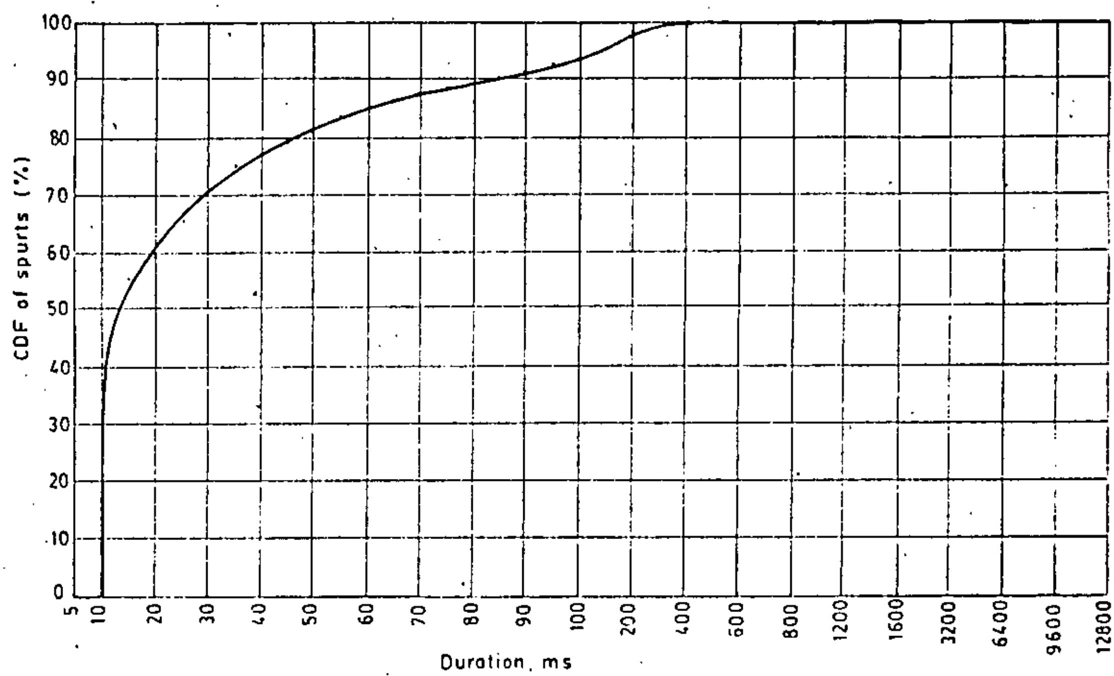


Figure 4.18 Overall CDF of spurts with a 10 ms hangover.

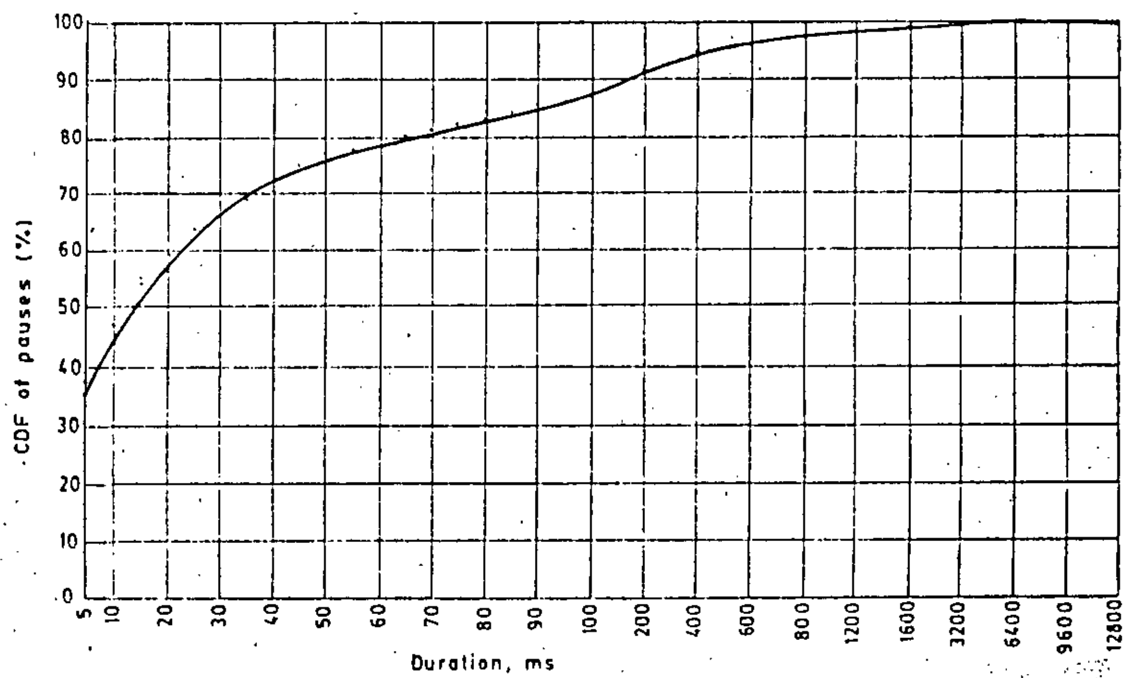


Figure 4.19 Overall CDF of pauses with a 10 ms hangover.

4.6 The effect of fill-in and hangover on different speech temporal parameters

To study the effect of fill-in on different speech temporal parameters the fill-in times of 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 150 and 200 ms are applied on the original data of every subject. For a particular fill-in time, the average values of speech temporal parameters are calculated using the data of 10 males and 10 females combined as a whole. The effect of fill-in on different speech temporal parameters are shown in table 4.16. The same is shown graphically in figures 4.20, 4.21, 4.22 and 4.23.

In a similar manner the effect of hangover time on different speech tempooral parameters are calculated. The results are shown in table 4.17 and in figures 4.20, 4.21, 4.22 and 4.23.

In the next chapter some statistical models are developed which represent the on-off characteristics of conversational Bangla speech over telephone channel.

Speech parameters	Fill-in time												
	00 ms	10 ms	15 ms	20 ms	25 ms	30 ms	40 ms	50 ms	60 ms	80 ms	100 ms	150 ms	200 ms
Speech activity (%)	16.40	22.11	26.39	27.76	29.41	30.25	32.92	34.48	37.08	40.01	42.92	48.01	50.95
Average spurt time (ms)	09.59	27.98	52.29	63.61	79.45	88.42	122.92	147.05	196.99	271.31	370.16	643.62	884.75
Average pause time (ms)	48.86	98.63	145.93	165.67	190.84	204.03	250.65	279.70	334.98	407.90	494.02	700.77	858.02
Talkspurt rate (per min.)	1026.50	474.04	302.80	261.78	222.10	205.29	160.71	140.70	112.92	88.48	69.56	44.75	34.55

Table 4.16 **Variation of overall speech temporal parameters with fill-in time**

Speech parameters	Hangover time												
	00 ms	10 ms	15 ms	20 ms	25 ms	30 ms	40 ms	50 ms	60 ms	80 ms	100 ms	150 ms	200 ms
Speech activity (%)	16.40	30.00	33.95	36.48	38.66	40.51	43.63	46.20	48.35	51.77	54.47	59.14	62.39
Average spurt time (ms)	09.59	37.98	67.28	83.60	104.43	118.40	162.88	196.99	256.87	351.09	469.82	792.83	1083.26
Average pause time (ms)	48.86	88.63	130.93	144.54	165.84	174.03	210.65	229.70	274.98	327.90	394.02	550.77	658.02
Talkspurt rate (per min.)	1026.50	474.02	302.80	261.78	222.10	205.26	160.71	140.70	112.92	88.48	69.56	44.75	34.55

Table 4.17 Variation of overall speech temporal parameters with hangover time

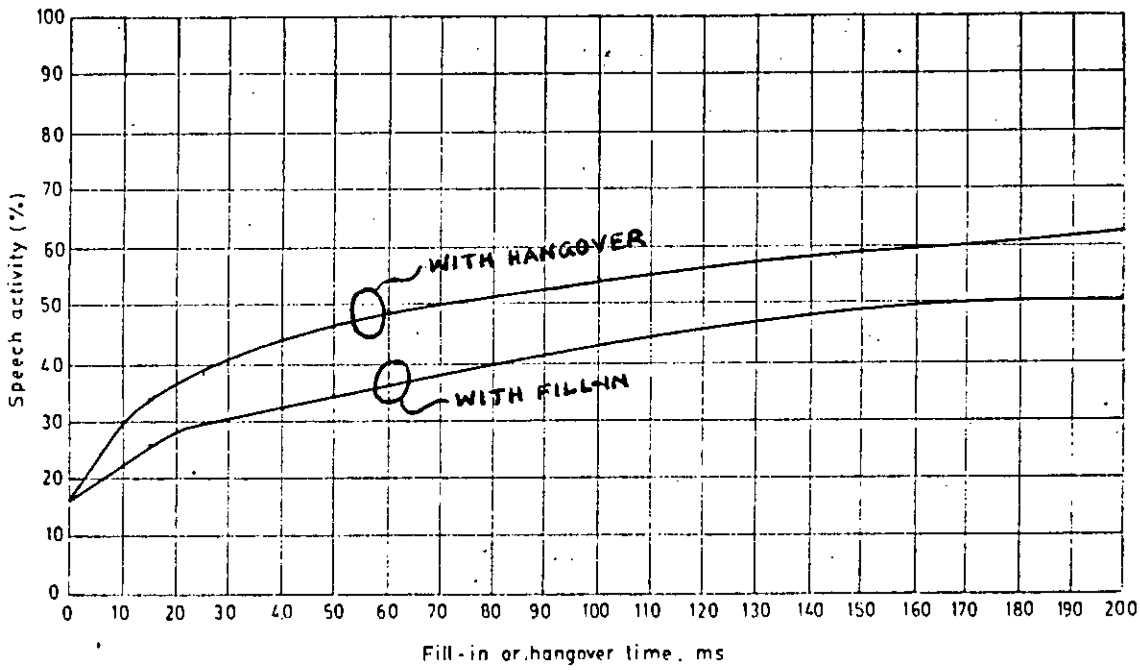


Figure 4.20 Speech activity vs. fill-in or hangover time.

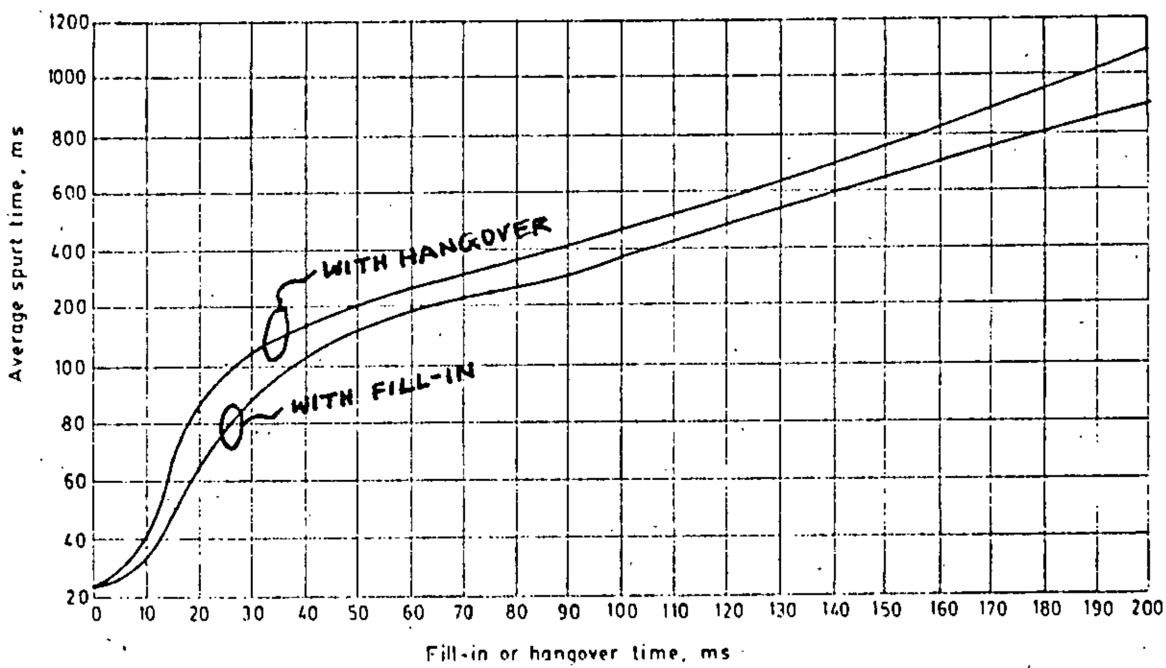


Figure 4.21 Average spurt time vs. fill-in or hangover time.

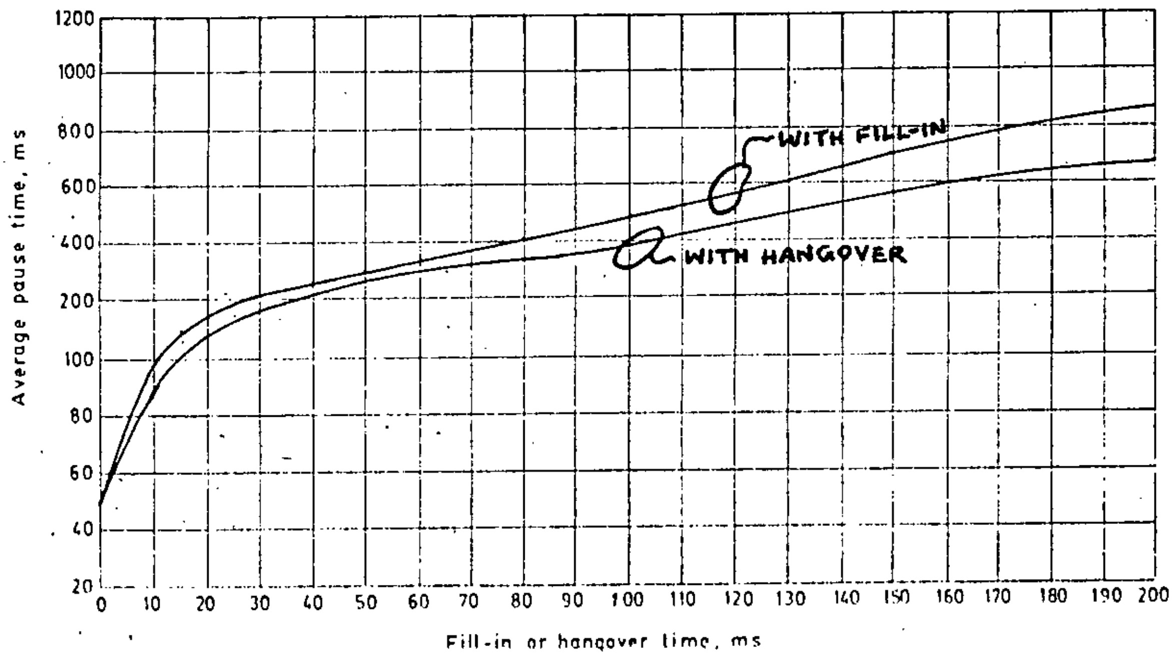


Figure 4.22 Average pause time vs. fill-in or hangover time.

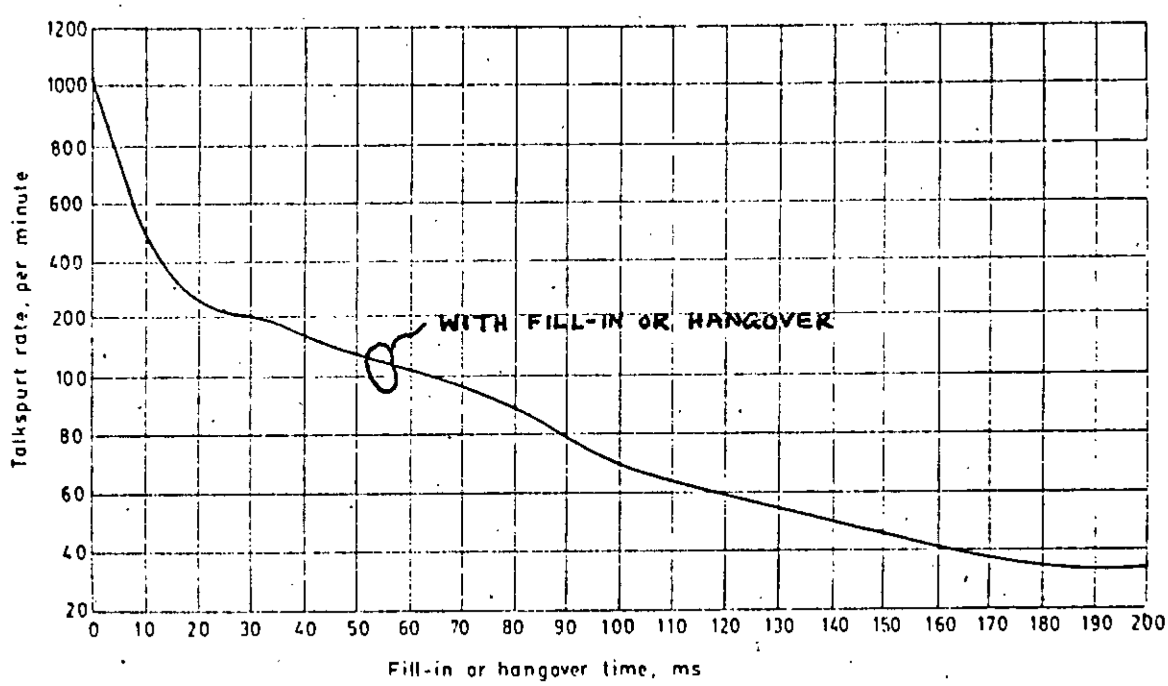


Figure 4.23 Talkspurt rate vs. fill-in or hangover time.

Chapter 5

Statistical modeling of on-off characteristics of conversational Bangla speech

5.1 Introduction

We have already calculated the overall cumulative distribution functions of spurts and pauses and different speech temporal parameters of conversational Bangla speech from the experimental data. In this chapter an attempt has been made to express the on-off characteristics in terms of their mathematical models. This would enable one to calculate the required characteristics of conversational Bangla speech from the suggested mathematical models.

5.2 The choice of the model

Different approximate methods for linear and non-linear curve-fitting are used in various fields of applied science. The visual inspection of the curves of figures 4.12, 4.13, 4.18, 4.19, 4.20, 4.21, 4.22 and 4.23 shown in chapter 4 representing the characteristics of conversational Bangla speech reveals the fact that they are all non-linear in nature. The generally used non-linear functions are:

- a power function of the form

$$f = b_1 t^{b} \qquad \ldots \quad \ldots \quad (5.1)$$

- an exponential function of the form

$$f = c_1 \exp(c_2 t) \qquad \ldots \quad \ldots \quad (5.2)$$

- a polynomial of the form

$$f = a_0 + a_1 t + a_2 t^2 + \dots \quad \dots + a_m t^m \qquad \dots \quad \dots \quad (5.3)$$

where,

f is a function of t and b_1, b_2, c_1, c_2, a_0, a_1, a_2, ... a_m etc. are the arbitrary constants and m is the degree of the polynomial.

The method of least squares is widely used to fit a set of data to the aforementioned non-linear curves. Our experimental data fits a polynomial the best. A generalized computer program named 'CFIT.FOR' (Appendix D) is prepared in FORTRAN-77 for fitting a polynomial by the method of least squares. This program can fit a polynomial of upto ninth degree which serves our purpose.

5.2.1 Mathematical models of the overall cumulative distribution functions of talkspurts and pauses

The nature of the curves shown in figures 4.12, 4.13, 4.18, and 4.19 of chapter 4 representing the overall cumulative distribution functions of talkspurts and pauses appears to us at the first inspection, to be exponential in nature. Therefore, an attempt has been made to fit the experimental data to the exponential function represented by equation (5.2).But the fit is not up to the expectation. In view of achieving a better fit next attempt has been made to fit the data to the power function represented by equation (5.1). This time again the result does not show any marked improvement. Finally, the experimental data is fitted to a polynomial function represented by equation (5.3) and it appears to be the best fit. Therefore, the polynomial functions are chosen to represent the overall cumulative distribution functions of talkspurts and pauses.

It should be mentioned here that the method of trial and error is used to choose the degree of the polynomial that gives the best fit. It is noted that higher the degree of the polynomial, the better is the result. But a serious drawback of using a higher order polynomial is that it shows oscillations of the data at higher values of time [Table 5.1] and it shows some CDF values greater than 100% which is practically impossible. This is because most of the spurts and pauses have lower durations and their change in number is also high at the lower values of time. On the otherhand, spurts and pauses having longer durations are less and at the higher values of time the change in number is not significant. It then appears that the curve representing the CDF of spurts or pauses has two distinct regions: one at the lower values of time showing higher rate of change and the other at the higher values of time showing a negligible rate of change.

This means a single curve may not fit such a set of data. In order to ignore the oscillations and at the same time to achieve a better fit, a two-segment fit is suggested: one for the lower values of time and the other for the higher values.

Table 5.1 Overall CDF of spurts with a 10 ms fill-in

Time in ms	Actual CDF values (%)	CDF values from the model (%)
5	55.71	70.63
10	59.87	71.96
15	66.04	73.27
20	70.25	74.54
25	75.19	75.80
30	77.60	77.03
35	79.90	78.23
40	81.66	79.41
45	83.99	80.56
50	85.33	81.70
55	86.53	82.80
60	87.67	83.89
65	88.94	84.95
70	89.85	85.99
75	90.76	87.00
80	91.55	87.99
85	92.37	88.96
90	92.91	89.91
95	93.46	90.84
100	93.90	91.74
200	98.38	105.61
400	99.71	114.20
800	99.97	93.66
1600	100.00	100.76
3200	100.00	99.97

The fitted polynomial function is given by

$$fsf = 69.2828 + .273614\,t - .520541 \times 10^{-3}\,t^{2} + .315404 \times 10^{-6}\,t^{3} - .557871 \times 10^{-10}\,t^{4} \quad \ldots \quad (5.4)$$

Where,
fsf= Overall CDF of spurts in % and
t= time in ms.

The results thus obtained show a marked improvement over the single curve and there is basically no difference between the fitted data and the experimental data.

The mathematical models for overall CDFs of spurts and pauses with a 10 ms fill-in time are represented by equations (5.5) and (5.6) respectively and those with a 10 ms hangover time are represented by equations (5.7) and (5.8) respectively. They are also shown graphically in figures 5.1, 5.2, 5.3 and 5.4. The plots of the experimental data are shown on the same graphs for comparision.

The mathematical models for CDFs of spurts and pauses are presented below.

<u>CDF of spurts with a 10 ms fill-in (fsf)</u>

$$fsf = 47.7949 + 1.54331\ t - .0238543\ t^{2} + .190553 \times 10^{-3}\ t^{3} - .601838 \times 10^{-6}\ t^{4} \qquad \ldots(5.5a)$$

$$\text{for} \quad 5\text{ms} \leqslant t \geqslant 95\ \text{ms};$$

$$fsf = 84.8524 + .119232\ t - .319281 \times 10^{-3}\ t^{2} + .327984 \times 10^{-6}\ t^{3} - .107068 \times 10^{-9}\ t^{4} \qquad \ldots(5.5b)$$

$$\text{for } 100\ \text{ms} \leqslant t \geqslant 1600\ \text{ms};$$

Where,

fsf represents the CDF of spurts in percent with a 10 ms fill-in and

t represents the duration of spurts in ms.

The equations are valid for the mentioned limits of time. The CDF value for

$t=0$ is zero and experimental data show that for values of $t > 1600$ ms the CDF value is always 100%.

<u>CDF of pauses with a 10 ms fill-in (fpf)</u>

$$fpf = -2.35929 + 3.32683\ t - .0568448\ t^{2} + .478436 \times 10^{-3}\ t^{3} - .155047 \times 10^{-5}\ t^{4} \qquad \ldots(5.6a)$$

$$\text{for } 15 \text{ ms} \leqslant t \geqslant 95 \text{ ms};$$

$$fpf = 79.4815 + .077823\ t - .109823 \times 10^{-3}\ t^{2} + .586701 \times 10^{-7}\ t^{3} - .979127 \times 10^{-11}\ t^{4} \qquad \ldots(5.6b)$$

$$\text{for } 100 \text{ ms} \leqslant t \geqslant 3200 \text{ ms};$$

where,

fpf represents the CDF of pauses in percent with a 10 ms fill-in and

t represents the duration of pauses in ms.

The equations are valid for the mentioned limits of time. The CDF value for $t \leqslant 10$ ms is zero and experimental data show that for values of $t > 3200$ ms the CDF value is always 100%.

CDF of spurts with a 10 ms hangover (fsh)

$$fsh = 30.3863 + 2.016\ t - .027711\ t^2 + .184322 \times 10^{-3}\ t^3 - .461347 \times 10^{-6}\ t^4 \quad \ldots(5.7a)$$

for $15\ ms \leqslant t \geqslant 95\ ms$;

$$fsh = 82.2474 + .140653\ t - .37885 \times 10^{-3}\ t^2 + .388832 \times 10^{-6}\ t^3 - .127039 \times 10^{-9}\ t^4 \quad \ldots(5.7b)$$

for $100\ ms \leqslant t \geqslant 1600\ ms$;

Where,

fsh represents the CDF of spurts in percent with a 10 ms hangover and

t represents the duration of spurts in ms.

The equations are valid for the mentioned limits of time. The CDF value for $t \leqslant 10$ ms is zero and experimental data show that for values of $t > 1600$ ms the CDF value is always 100%.

CDF of pauses with a 10 ms hangover (fph)

$$fph = 25.9017 + 2.28631\, t - .0414111\, t^{2} + .381296 \times 10^{-3}\, t^{3} - .135216 \times 10^{-5}\, t^{4} \quad \ldots(5.8a)$$

$$\text{for} \quad 5\ ms \leqslant t \geqslant 95\ ms;$$

$$fph = 85.3405 + .0328155\, t - .247857 \times 10^{-4}\, t^{2} + .683572 \times 10^{-8}\, t^{3} - .579413 \times 10^{-12}\, t^{4} \quad \ldots(5.8b)$$

$$\text{for } 100\ ms \leqslant t \geqslant 6400\ ms;$$

Where,

fph represents the CDF of pauses in percent with a 10 ms hangover and

t represents the duration of pauses in ms.

The equations are valid for the mentioned limits of time. The CDF value for $t=0$ is zero and experimental data show that for values of $t > 6400$ ms the CDF value is always 100%.

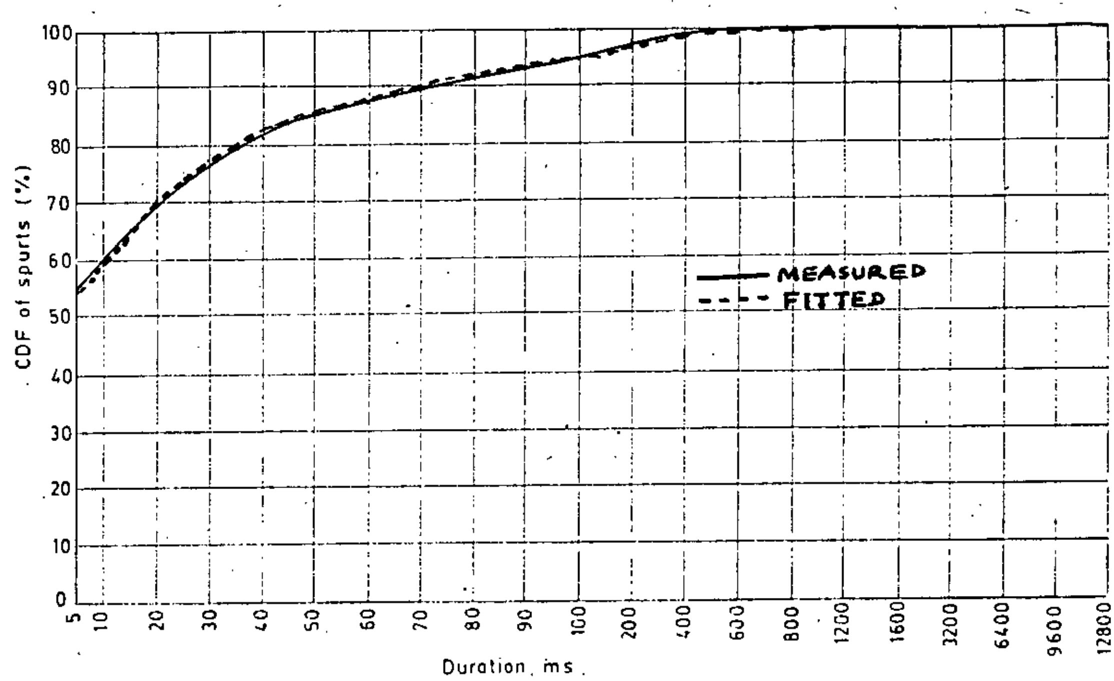


Figure 5.1 Overall CDF of spurts with a 10 ms fill-in time.

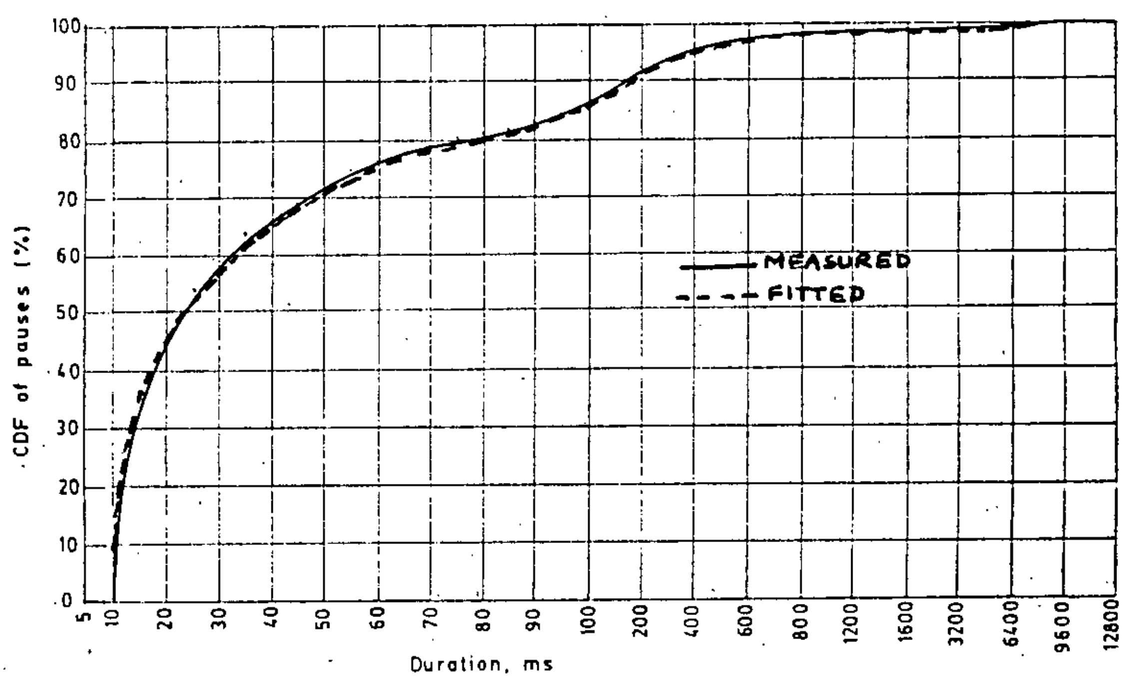


Figure 5.2 Overall CDF of pauses with a 10 ms fill-in time.

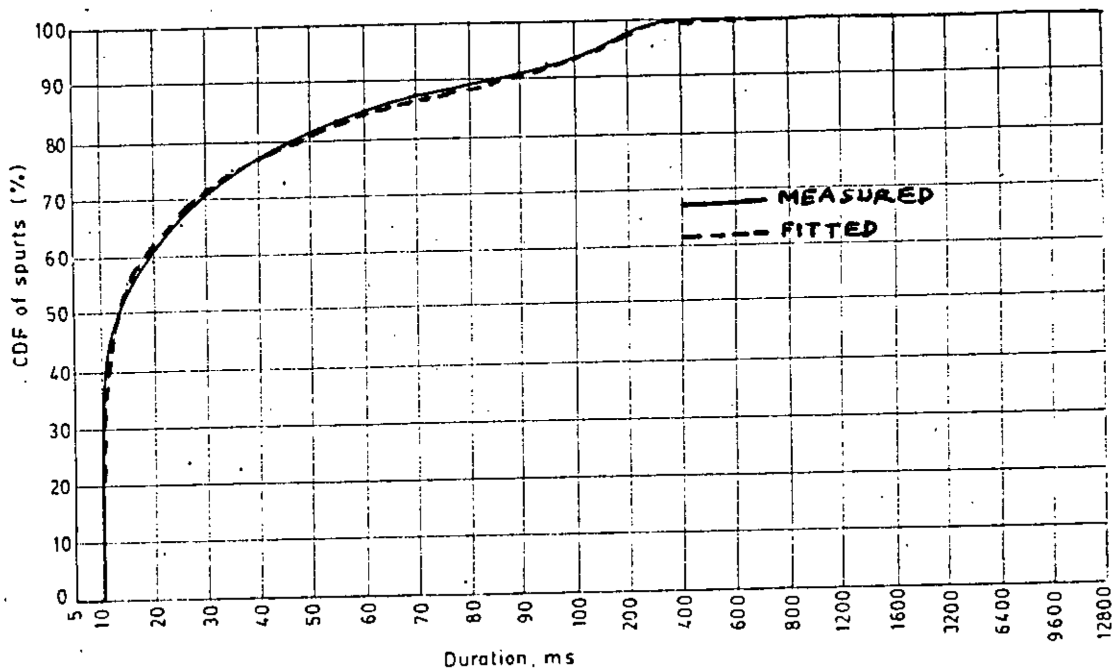


Figure 5.3 Overall CDF of spurts with a 10 ms hangover time.

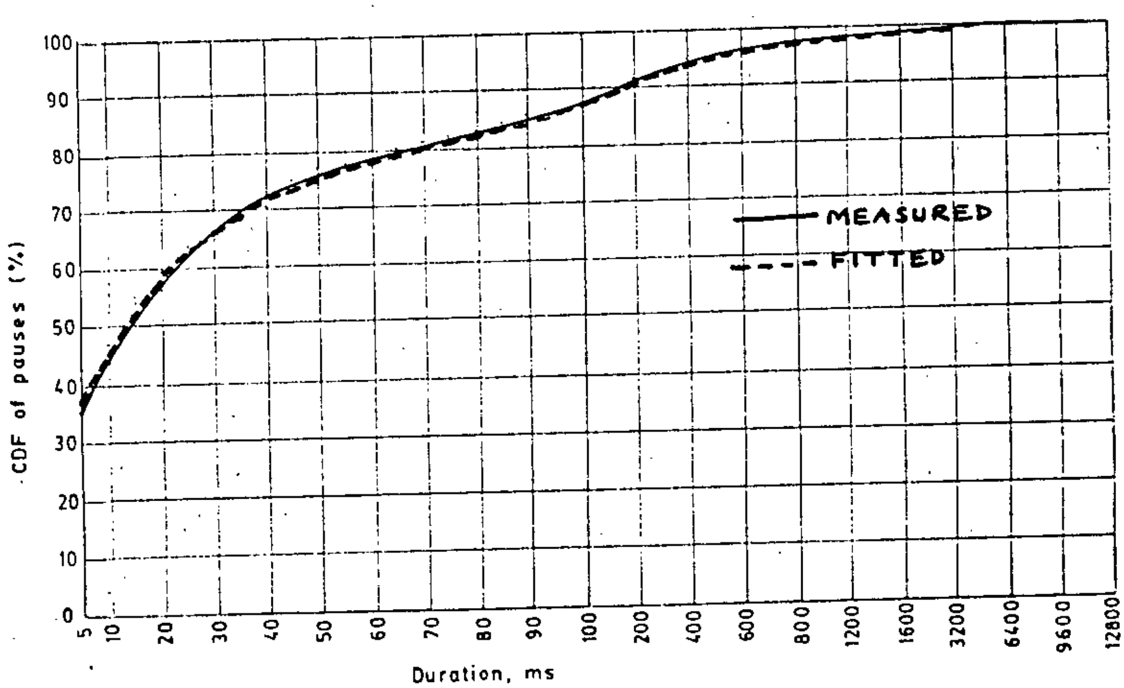


Figure 5.4 Overall CDF of pauses with a 10 ms hangover time.

5.2.2 Mathematical models of the speech temporal parameters as a function of fill-in or hangover time

The curves of figures 4.20, 4.21, 4.22 and 4.23 as shown in chapter 4 representing the different speech temporal parameters of conversational Bangla speech are also non-linear in nature. The experimental data for these speech parameters have also given the best to the polynomial functions. The method of trial and error has been followed to choose the degree of the polynomial to achieve the best fit. Unlike the CDFs, each speech temporal parameter is fitted by a single polynomial. The mathematical models for speech activity, average spurt time, average pause time and talkspurt rate as a function of fill-in time are given by the equations (5.9), (5.10), (5.11) and (5.12) respectively and the same for those as a function of hangover time are given by the equations (5.13), (5.14), (5.15) and (5.12) respectively. The variation of talkspurt rate with fill-in time and with hangover time are exactly the same and it is represented by equation (5.12). They are also shown graphically in figures 5.5, 5.6, 5.7 and 5.8 respectively. The effect of fill-in and hangover on speech parameters are shown on the same graph.

The mathematical models for different speech temporal parameters as a function of fill-in or hangover time are presented below. All the equations are valid for a practical range of fill-in or hangover time of about 250 ms.

<u>Speech activity as a function of fill-in time</u>

$$saf = 16.9951 + .633731\ t - .721589 \times 10^{-2}\ t^{2} + .434876 \times 10^{-4}\ t^{3} - .950729 \times 10^{-7}\ t^{4} \qquad \ldots(5.9)$$

where,

saf= speech activity in percent and

t= fill-in time in ms.

<u>Average spurt time as a function of fill-in time</u>

$$astf = 8.3204 + 2.64665\, t - .129431 \times 10^{-2}\, t^{2} + .16775 \times 10^{-3}\, t^{3} - .589486 \times 10^{-6}\, t^{4} \qquad \ldots(5.10)$$

where,

astf= average spurt time in ms and

t= fill-in time in ms.

<u>Average pause time as a function of fill-in time</u>

$$aptf = 49.8959 + .0631451\, t - .0445028\, t^{2} + .340812 \times 10^{-3}\, t^{3} - .875931 \times 10^{-6}\, t^{4} \qquad \ldots(5.11)$$

where,

aptf= average pause time in ms and

t= fill-in time in ms.

Talkspurt rate as a function of fill-in or hangover time

$$trfh = 1028.41 - 88.7137\,t + 3.99793\,t^{2} - .0971569\,t^{3} + .132159 \times 10^{-2}\,t^{4} - .100021 \times 10^{-4}\,t^{5} + .390263 \times 10^{-7}\,t^{6} - .607389 \times 10^{-10}\,t^{7} \quad \ldots(5.12)$$

where,

trfh= talkspurt rate per minute and

t= fill-in time or hangover time in ms.

Speech activity as a function of hangover time

$$sah = 18.5003 + 1.12248\,t - .0150264\,t^{2} + .925183 \times 10^{-4}\,t^{3} - .19989 \times 10^{-6}\,t^{4} \quad \ldots(5.13)$$

where,

sah= speech activity in percent and

t= hangover time in ms.

Average spurt time as a function of hangover time

$$asth = 8.31545 + 3.64755\,t - .134686 \times 10^{-2}\,t^{2} + .167895 \times 10^{-3}\,t^{3} - .589938 \times 10^{-6}\,t^{4} \quad \ldots(5.14)$$

where,

asth= average spurt time in ms and

t= hangover time in ms.

<u>Average pause time as a function of hangover time</u>

$$apth= 49.8294 + 5.30215\ t - .0441341\ t^{2} + .33777 \times 10^{-3}\ t^{3} - .868342 \times 10^{-6}\ t^{4} \quad ...(5.15)$$

where,

apth= average pause time in ms and

t= hangover time in ms.

Thus we have developed the mathematical models for different on-off characteristics of conversational Bangla speech and one can now easily calculate the required characteristic values for conversational Bangla speech using the mathematical models represented by the equations (5.5) to (5.15).

The next chapter deals with the experimental results and discussions of our analysis.

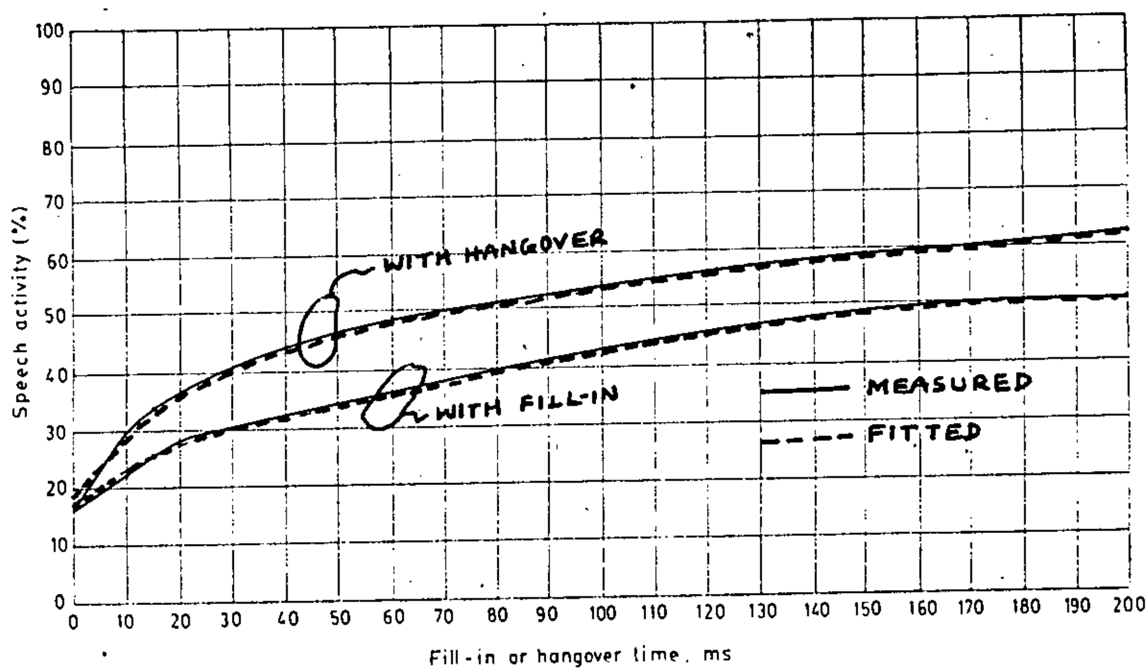


Figure 5.5 Speech activity vs. fill-in or hangover time.

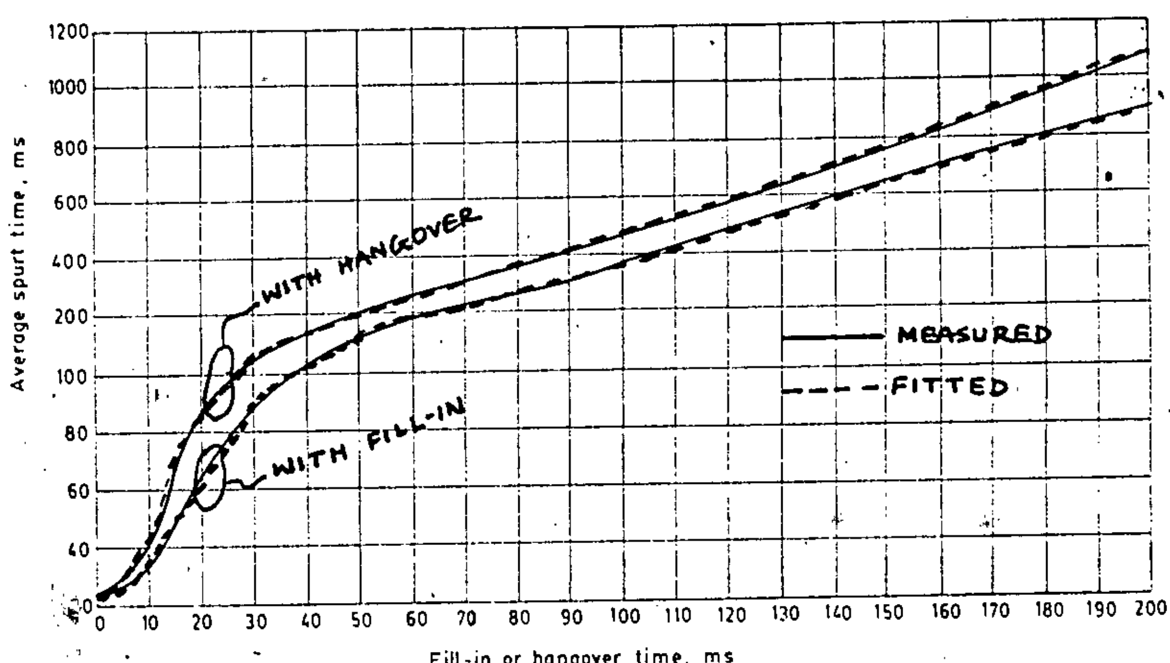


Figure 5.6 Average spurt time vs. fill-in or hangover time.

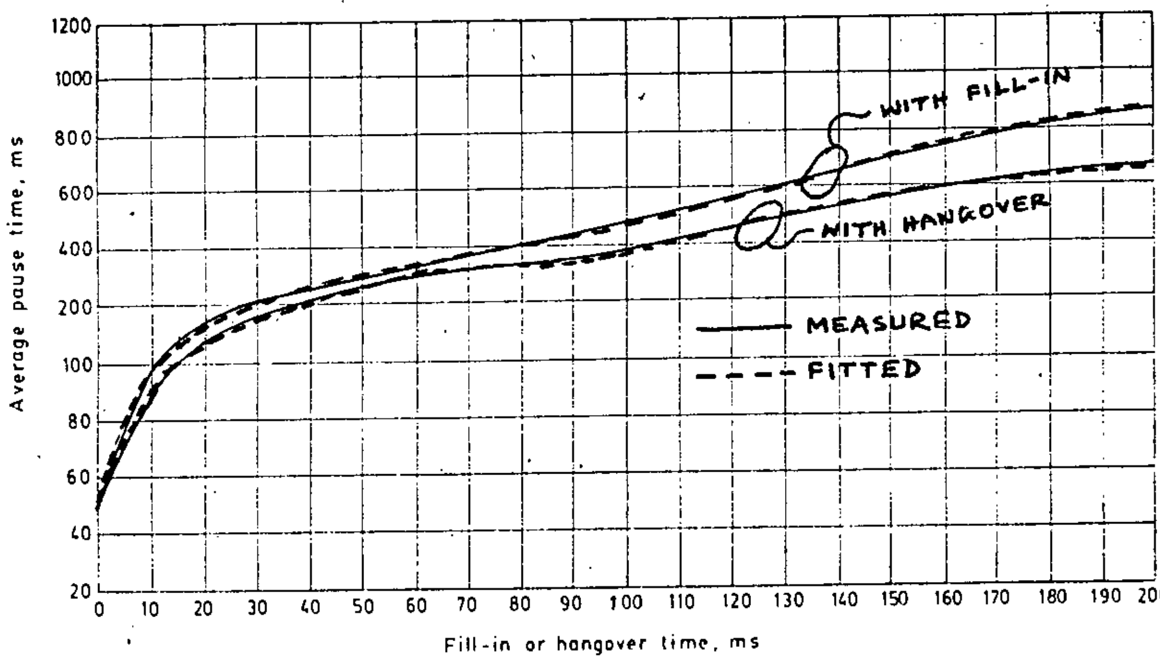


Figure 5.7 Average pause time vs. fill-in or hangover time.

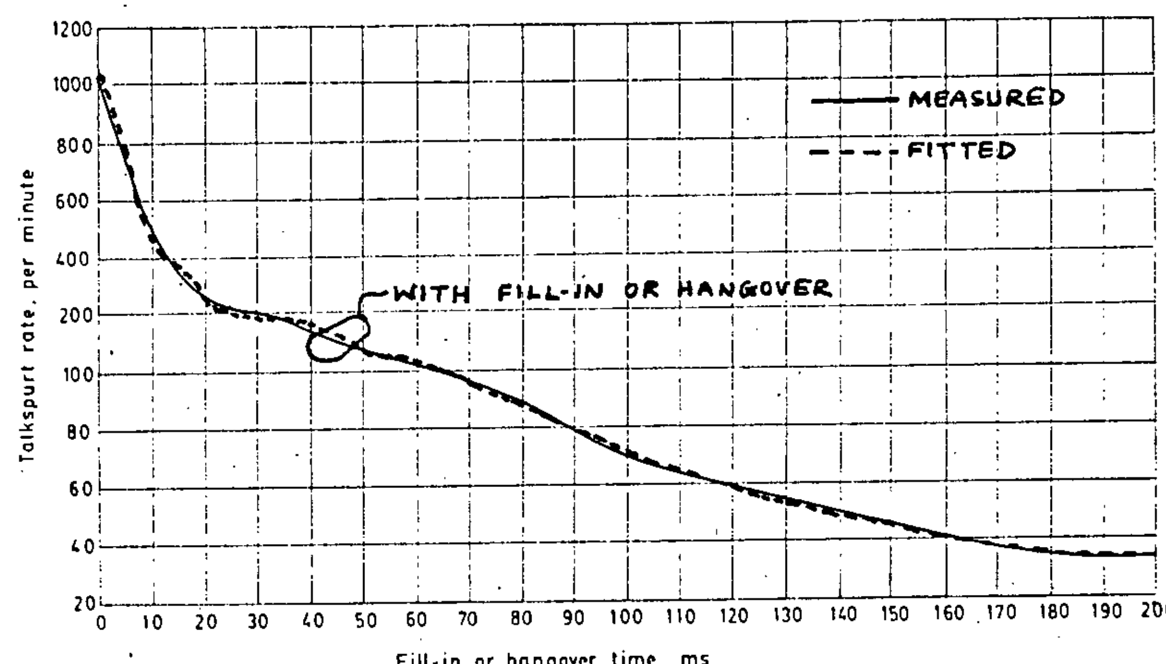


Figure 5.8 Talkspurt rate vs. fill-in or hangover time.

Chapter 6

Experimental results and discussions

6.1 Introduction

In previous chapters the on-off patterns of Bangla speech are analysed using generated data-base. The speech temporal parameters are characterised and cumulative distribution functions are plotted with or without hangover or fill-in.

The present chapter elaborates the results and discusses the analysis that has so far been carried out.

6.2 On-off characteristics of conversational Bangla speech without fill-in or hangover

As already stated in chapter 4, the on-off characteristics calculated from the original data do not represent the true values of conversational Bangla speech. They are calculated just to see the limitations of the speech detector by noting the extent of deviation of speech parameter values from the true ones.

From table 4.3 it is seen that the speech activity varies from 8.63% to 26.59% for male voice. The average speech activity calculated from the data of 10 male speakers combined as a whole is 16.63%. The same table shows that average spurt time varies from 5.92 ms to 22.01 ms and the average spurt time for 10 male speakers is 11.21 ms. Average pause time varies from 24.90 ms to 125.27 ms and for 10 male speakers as a whole it is 56.21 ms. The talkspurt rate varies from 439.5 per minute to 1796 per minute and the average talkspurt rate is 389.92 per minute.

On the otherhand, from table 4.4 it is seen that the speech activity for female voice varies from 9.08% to 25.90%. The average speech activity calculated from the data of 10 female speakers combined as a whole is 16.18%. The average spurt time varies from 6.73 ms to 10.96 ms and the average spurt time for 10 female speakers is 8.34 ms. The average pause time varies from 19.27 ms to 78.48 ms and for 10 female speakers as a whole it is 43.25 ms. The talkspurt rate varies from 695.1 per minute to 2307.3 per minute and the average talkspurt rate is 1163.09 per minute.

Comparision of the speech temporal parameter values for male speakers with those for female speakers show that speech activities for male and female speakers are nearly equal. The average talkspurt rate of female speakers is much higher than that of male speakers. The average spurt time and the average pause time for female speakers are less than those for male speakers.

The CDF of spurts as a function of duration of time for male voice is shown in figure 4.2 and that for female voice is shown in figure 4.3. The observation of the curves show that upto 100 ms the CDF valuse of spurts for female voice are greater than those for male voice and above 200 ms the result is reversed. The CDFs of pauses as a function of duration of time for male and female voices are shown in figures 4.4 and 4.5 respectively. It is seen from these curves that upto 200 ms the CDF values for female voice are less than those for male voice and above 200 ms the result is reversed.

The overall speech temporal parameters taking the data of male and female speakers into account is shown in table 4.5. The overall speech activity is 16.40%, the average spurt time is 9.59 ms, the average pause time is 48.86 ms and the talkspurt rate is 1026.50 per minute.

The overall CDF of spurts is shown in figure 4.6 and that of pauses is shown in figure 4.7. Both the curves have non-zero initial values and after increasing gradually both of them become saturated.

6.3 On-off characteristics of conversational Bangla speech with a 10 ms fill-in

Speech temporal parameters calculated from the data after applying a 10 ms fill-in represent the true on-off characteristics of conversational Bangla speech. The intersyllabic gaps are no longer present in the data. They are considered here as part of the continuous speech segment. Table 4.8 lists the speech temporal parameters for male voice. It is seen from the table that the speech activity varies from 10.81% to 37.29% for male voice and the average speech activity is 22.02%. The average spurt time varies from 11.66 ms to 86.93 ms and the average spurt time for 10 male speakers is 35.85 ms. The average pause time varies from 48.15 ms to 302.01 ms and the average of 10 male speakers is 126.99 ms. The talkspurt rate varies from 154.6 per minute to 803.9 per minute and the average talkspurt rate is 368.54 per minute.

On the otherhand, table 4.9 shows that for female voice the speech activity varies from 11.95% to 39.61% and the average speech activity is 22.19%. The average spurt time varies from 14.87 ms to 27.58 ms and the average spurt time for 10 female speakers is 22.98 ms. The avaerage pause time varies from 42.06 ms to 138.92 ms and the overall value is 80.58 ms. The talkspurt rate varies from 373.9 per minute to 921.1 per minute and the average talkspurt rate is 579.54 per minute.

Comparision of the speech temporal parameter values for male speakers with those for female speakers show that speech activities for male and female speakers are nearly equal. The average talkspurt rate for female speakers is higher than that for male speakers. The average spurt time and the average pause time for female speakers are less than those for male speakers. All the aforementioned data show that on an average, female speakers talk more than male speakers in a given time.

The CDF of spurts as a function of duration of time for male voice is shown in figure 4.8. and for female voice is shown in figure 4.9. It is seen from the curves that female speakers have a higher initial CDF value and upto 200 ms the CDF values of spurts for female speakers is greater than those of male speakers. The result is reversed above 200 ms. It indicates that shorter spurts are higher in number for female speakers. The CDFs of pauses as a function of duration of time for male and female speakers are shown in figures 4.10 and 4.11 respectively. It is seen from the curves that upto 30 ms the CDF values of pauses for female speakers are less than those for male spekers and above 30 ms the result is reversed. It indicates that the number of pauses having durations of 30 ms and less are relatively higher in male spekers than in female speakers.

The overall CDF of spurts is shown in figure 4.12 and that of pauses is shown in Figure 4.13. The overall CDF of spurts shows higher values at the very beginning indicating the presence of higher number of shorter spurts. The curve shows that about 55.71% of the spurts have a 5 ms duration and most of the spurts have the durations of less then 200 ms. On the otherhand, at higher values of time the curve becomes saturated. It indicates that the spurts of longer durations are less in number. Similar is the case with the pause distribution. But the curve for pause distribution is shifted to the right indicating the absence of pauses having durations less than or equal to 10 ms. The curve has a zero initial value and the rate of rise is slower than that of spurt distribution indicating nearly an uniform distribution of pauses.

However, the overall speech temporal parameters listed in table 4.10 shows that average speech activity with a 10 ms fill-in is 22.11%. The average spurt time is 27.98 ms, the average pause time is 98.63 ms and the average talkspurt rate is 474.04 per minute. It is already stated that these represent the true speech temporal parameters for conversational Bangla speech over telephone channel.

When the speech temporal parameters without fill-in are compared to those with a 10 ms fill-in, it becomes evident that speech activity has increased from 16.40% to 22.11%. The increase in speech activity is the result of increased spurt time due to fill-in. The increase in speech activity by 5.71% may be considered equal to the percent of time comprising the intersyllabic gaps. The average spurt time has increased from 9.59 ms to 27.98 ms. This is because with fill-in the total number of spurts are decreased and at the same time the lengths of the spurts are increased. The average pause time has also increased from 48.86 ms to 98.63 ms. The reason is that, with fill-in, the number of pauses is decreased and the pauses less than or equal to the fill-in time are vanished leaving the longer pauses only. The talkspurt rate is sharply decreased from 1026.5 per minute to 474.04 per minute because of a decreased number of spurts.

6.4 On-off characteristics of conversational Bangla speech with a 10 ms hangover

The conventional detectors have no other way but to employ a slow release or delay to bridge the small intersyllabic gaps. Therefore, the on-off characteristics that are calculated from the data modified by a 10 ms hangover time corresponds to those obtainable from a conventional detector having the same hangover time of 10 ms.

Table 4.13 lists the speech temporal parameters with a 10 ms hangover for male voice. It is seen from the table that the speech activity varies from 14.38% to 47.72% for male voice.. The average speech activity is 28.16%. The average spurt time varies from 21.66 ms to 96.91 ms and the average spurt time for all the 10 male speakers is 45.85 ms. The average pause time varies from 38.15 ms to 292.01 ms and the average for 10 male speakers is 116.99 ms. The talkspurt rate varies from 154.6 per minute to 803.9 per minute and the average talkspurt rate is 368.54 per minute.

On the otherhand, table 4.14 shows that the speech activity for female speakers varies from 18.91% to 53.97% and the average speech activity is 31.85%. The average spurt time varies from 24.86 ms to 37.58 ms and the average spurt time for 10 female speakers is 32.98 ms. The average pause time varies from 32.06 ms to 128.92 ms and the average of 10 female speakers is 70.59 ms. The talkspurt rate varies from 373.9 per minute to 921.1 per minute and the average talkspurt rate is 579.51 per minute.

Comparision of the speech temporal parameter values for male speakers with those for female speakers show that speech activity for female speakers is higher in this case than that for male speakers. This has already been noted that number of talkspurts or pauses in a female voice are higher than those in a male voice. And when hangover or delay is applied to bridge the in-tersyllabic gaps, the beginning of all the other gaps greater than hangover time are also filled in by the delay time. This results in an increased spurt time. As the number of pauses are higher in female voice than that in male voice, the delay contributes a higher summation of spurt durations in female speech and hence the increased speech activity. The average spurt time and the average pause time for female speakers are less than those for male speakers. The talkspurt rate for female speakers is higher than that for male speakers.

The CDF of spurts as a function of duration of time with a 10 ms hang-over for male voice is shown in figure 4.14 and for female voice is shown in figure 4.15. It is seen that upto 200 ms the CDF values for female voice are greater than those for male voice and above 200 ms the result is reversed. The CDF of pauses as a function of duration of time with a 10 ms hangover time for male voice is shown in figure 4.16 and that for female voice is shown in figure 4.17. It is seen from the curves that upto 20 ms the CDF values of pauses for female voice are less than those for male voice and above 20 ms the result is reversed.

The overall CDF of spurts with a 10 ms hangover is shown in figure 4.18 and that of pauses is shown in figure 4.19. The curve for CDF of spurts has been shifted to the right showing absence of spurts that are less than or equal to 10 ms and the curve for CDF of pauses has been shifted to the left showing 36.13% presence of pauses having 5 ms duration.

However, the overall speech temporal parameters with a 10 ms hangover are listed in table 4.15. The overall speech activity is 30%, the average spurt time is 37.98 ms, the average pause time is 98.63 ms and the talkspurt rate is 474.02 per minute.

Comparision of the speech temporal parameters with a 10 ms fill-in with those with a 10 ms hangover shows that speech activity has increased from 22.11% to 30%. The increase is due to the effect of delay. The applied delay fills in the intersyllabic gaps as well as the beginning of every gap which are greater than hangover or delay time. Therefore, it is noted that in a conventional detector, to fill in the intersyllabic gap of 5.71% an additional 7.89% (i.e. (30-22.11)%) of the time is lost due to the applied hangover.

The average spurt time has increased from 27.98 ms to 37.98 ms due to the increase in spurt duration and decrease in the number of spurts. But the average pause time has reduced from 98.63 ms to 88.63 ms. It is because with hangover, the number of pause as well as the duration of pauses are reduced.

6.5 Variation of speech temporal parameters with fill-in or hangover

Table 4.16 shows the variation of different speech temporal parameters with fill-in time and table 4.17 shows the variation of the same with hangover time.

Figure 4.20 shows the variation of speech activity with fill-in or hangover time. It is noted that speech activity increases with increase in hangover or fill-in time. The curve with hangover lies above the one with fill-in. The speech activity varies from 16.40% with zero fill-in or hangover to 50.95% with a 200 ms fill-in and to 62.39% with a 200 ms hangover.

Figure 4.21 shows the variation of average spurt time as a function of fill-in or hangover time. It is seen from the curves that the average spurt time increases with increase in fill-in or hangover time. But the curve with hangover lies above the one with fill-in. The average spurt time varies from 9.59 ms with zero fill-in or hangover to 884.75 ms with a 200 ms fill-in and to 1083.26 ms with a 200 ms hangover.

Figure 4.22 shows the variation of average pause time as a function of fill-in or hangover time. It is seen from the curves that the average pause time increases with fill-in or hangover time. But the curve with hangover lies below the one with fill-in. The average pause time varies from 48.86 ms with zero fill-in or hangover to 858.02 ms with a 200 ms fill-in and to 658.02 ms with a 200 ms hangover.

Figure 4.23 shows the variation of talkspurt rate with fill-in or hangover. The curve shows that the talkspurt rate decreases with increase in fill-in or hangover time. However, it is noted that the variation with fill-in and that with hangover are the same. The talkspurt rate varies from 1026.50 per minute with zero fill-in or hangover to 34.55 per minute with a 200 ms fill-in or hangover.

6.6 Comparision of the measured and fitted characteristics of conversational Bangla speech

The CDF of spurts and pauses with a 10 ms fill-in have been modeled in equations (5.5) and (5.6) respectively and those with a 10 ms hangover have been modeled in equations (5.7) and (5.8) respectively. To achieve a better accuracy, we have fitted each of them to the 4_{th} order polynomials in two segments. The calculated CDF values of talkspurts and pauses from their models agree well with the measured data.

The speech temporal parameters with a 10 ms fill-in and a 10 ms hangover are also modeled in equations (5.9) to (5.15). Each of the speech parameters shows better fit to polynomials of the 4_{th} order except the talkspurt rate. It has been fitted to a 7_{th} order polynomial and has been represented by equation (5.12). The measured values of speech temporal parameters have been compared to those calculated from the models and also shown to be in good agreement for a practical range of hangover and fill-in time.

Therefore, one can use the suggested mathematical models to find out the required on-off characteristics of conversational Bangla speech without any appreciable error.

6.7 Comparision of speech temporal parameters of Bangla and English

The on-off characteristics of English speech [2] are shown in graphs and tables in Appendix E. The comparision is given in tabular forms in this chapter.

Table 6.1 Speech parameters for Bangla and English with zero fill-in or hangover

Speech temporal parameters	Bangla speech	English speech
Speech activity	16.40%	27.50%
Average spurt time	9.59 ms	250.00 ms
Average pause time	48.86 ms	600.00 ms
Talkspurt rate	1026.50 per min.	72.00 per min.

Table 6.1 lists the various speech temporal parameters for Bangla and English speech with zero fill-in or hangover time. It is seen from the table that speech activity for Bangla speech is lower than that for English speech. The average spurt and pause time are much lower and the talkspurt rate is much higher.

Table 6.2 Speech parameters for Bangla and English with a 10 ms fill-in

Speech temporal parameters	Bangla speech	English speech
Speech activity	22.11%	27.70%
Average spurt time	27.98 ms	260.00 ms
Average pause time	98.63 ms	630.00 ms
Talkspurt rate	474.04 per min.	65.00 per min.

Table 6.2 lists the speech temporal parameters with a 10 ms fill-in. As already stated in section 4.5.2 , these values represent the true parameter values of conversational Bangla speech. The speech activity for Bangla speech has been found to be 22.11% , whereas that for English speech is 27.70%. The average spurt and pause time for Bangla speech are lower and the talkspurt rate is higher.

Table 6.3 Speech parameters for Bangla and English with a 10 ms hangover

Speech temporal parameters	Bangla speech	English speech
Speech activity	30.00%	28.50%
Average spurt time	37.98 ms	280.00 ms
Average pause time	88.63 ms	630.00 ms
Talkspurt rate	474.02 per min.	65.00 per min.

Table 6.3 shows the speech parameters with a 10 ms hangover time. These correspond to the values obtainable from a detector employing a delay of 10 ms. The result shows a higher speech activity, a lower average spurt time, a lower average pause time and a higher talkspurt rate of Bangla speech compared to English speech. The main reason for higher speech activity with hangover is the higher talkspurt rate of Bangla speech. With hangover, the initial portion of every pause greater than the hangover time is filled in by the hangover time. As Bangla speech has a higher talkspurt rate, it contains higher number of pauses compared to English speech. Therefore, the total spurt time of

Bangla speech is increased by the amount of time equal to the hangover time multiplied by the difference in number of pauses and hence an increased speech activity.

Table 6.4 Speech parameters for Bangla and English with a 200 ms fill-in

Speech temporal parameters	Bangla speech	English speech
Speech activity	50.95%	31.50%
Average spurt time	884.75 ms	1110.00 ms
Average pause time	858.02 ms	2400.00 ms
Talkspurt rate	34.55 per min.	16.67 per min.

Table 6.4 lists the speech parameter values with a 200 ms fill-in time. At this stage the intersyllabic gaps and most of the inter word gaps are filled in. The speech activity for Bangla speech is seen to have a value of 50.95%, a considerable increase. But the speech activity for English speech has increased to only 31.50%. This indicates that most of the pauses in English speech are greater than 200 ms. The average spurt time for Bangla speech approaches to that for English speech, but the average pause time still lags far behind. The talkspurt rate is nearly twice the talkspurt rate of English speech.

Table 6.5 Speech parameters for Bangla and English with a 200 ms hangover

Speech temporal parameters	Bangla speech	English speech
Speech activity	62.39%	37.00%
Average spurt time	1083.26 ms	1300.00 ms
Average pause time	658.02 ms	2220.00 ms
Talkspurt rate	34.55 per min.	16.67 per min.

Table 6.5 lists the speech parameters with a 200 ms hangover time. Conventional TASI detectors usually employ a hangover or delay of about 200 ms. The results show that the speech activity of Bangla speech is much higher than that of English speech. The average spurt time is nearly equal to the average spurt time of English speech but the average pause time lags far behind. The talkspurt rate is nearly twice the talkspurt rate of English speech. Though the true speech activity for Bangla speech is not greater than 30%, it shows such a high speech activity of 62.39% for the reasons already explained in the discussions under table 6.3.

Chapter 7

Conclusions and suggestions for further work

7.1 Conclusions

Speech interpolation technique is used in present day communication to ensure maximum utilization of the available communication channels. The measurement and modeling of on-off speech characteristics are essential for the design and performance analysis of speech interpolation systems. In this study the on-off characteristics of conversational Bangla speech over telephone channels are measured using a large group of data-base. Some mathematical models are also suggested for them.

Measurement and calculations of different on-off speech statistics from the suggested mathematical models have been compared. The calculated values of speech temporal parameters of Bangla have also been compared to those of English.

The main results of the study are summarized as follows.

The speech temporal parameter values calculated from the speech detector output without fill-in or hangover can not be used as the true speech parameter values. Because the deviation is much higher from the true values.

Though the average speech activity for male and female speakers with fill-in are nearly the same, the talkspurt rate for female speakers is much higher.

The average spurt time and the average pause time for female speakers are less than those for male speakers.

With hangover the speech activity for female speakers is higher than that for male speakers.

The models for CDFs of talkspurts and pauses with a 10 ms fill-in and a 10 ms hangover are good approximations as they agree well with the measured data.

Directly measured values of speech temporal parameters and those computed from the fitted mathematical models have been compared and shown to be in good agreement for a practical range of hangover and fill-in time.

According to our results, the talkspurt rate decreases from 1026.50 per minute with zero hangover time to 34.55 per minute with a 200 ms hangover time, while the speech activity with time increases considerably from 16.40 to about 62.39 percent. Therefore, the tradeoff between reduced talkspurt and increased speech activity with hangover should be given due consideration for conversational speech.

Bangla conversational speech shows lower speech activity,lower average spurt time and higher average pause time with fill-in than with hangover, and it shows the same talkspurt rate with fill-in or hangover.

When compared to the parameters of English speech over telephone channels, Bangla conversational speech shows a lower average spurt time, a lower average pause time and a higher talkspurt rate. As a result it shows higher speech activity with a fixed value of fill-in or hangover time compared to its English speech counterpart, while with a 10 ms fill-in or hangover it shows lower speech activity.

From the analysis of conversational Bangla speech over telephone channels it is understood that if a speech interpolation system is designed considering the on-off characteristics of speech of only one particular language , it may not show the same performance for other languages. As an example TASI detectors are designed by employing a hangover time of 200 to 250 ms for English speech showing a speech activity of about 40%. But the same hangover time of 200 ms, if employed for Bangla speech shows a speech activity of about 62.39%. The result is of no wonder as the true on-off patterns for Bangla speech are those obtained by employing a 10 ms fill-in time, which shows a speech activity of about 22.11%. It signifies that if a 200 ms hangover is employed for Bangla speech we are occupying the channel for about 62.39% of the total time and using only 22.11% of it. Therefore, a lower hangover time is suggested for Bangla speech to have effectively a lower speech activity and thereby achieve a higher trunk to channel ratio.

From the analysis of the on-off speech characteristics of conversational Bangla speech and their comparision with those of English speech it becomes evident that if a speech interpolation system has to be designed where people of different languages will talk over telephone channels of the system, due consideration should be given to the on-off patterns of all the languages.

7.2 Suggestions for further work

A number of useful extensions of the work described in this thesis are possible. They are discussed in the following paragraph.

This work suggests to employ a lower hangover time for Bangla speech than that for English speech so that effective speech activity becomes low resulting in a high trunk to channel ratio. But it should be noted that with a low hangover time the talkspurt rate of conversational Bangla speech increases. Therefore, the probability of speech clipping during busy hours becomes higher. Further study is required to optimize speech clipping by selecting a proper hangover time for the speech detector for Bangla speech. The present analysis and other design constraints of the speech detector may be of useful reference. A suggestion for avoiding speech clipping during busy hours is to incorporate delay capabilities into a DSI or TASI system. It has been stated in section 2.3.7 that when the number of active sources exceeds the number of channels in a TASI or DSI system, the usual operation is to block the overload and cause clipping of some speech syllables. To avoid speech clipping the blocked syllables may be stored in a buffer and delayed for a short period of time in anticipation that a channel will soon become available. The reconstructed speech from this process will give a better quality than when conventional syllable clipping occurs. Therefore, another study may be carried out for the feasibility of incorporating delay capabilities (temporary storage of speech signals in a buffer) for Bangla speech interpolation to enhance the greater levels of line utilization.

Appendix A

Characteristics of Bangla alphabet

Table III -Average amplitude, duration and bandwidth of Bangla alphabet in male voice

Letter of the alphabet	Amplitude V_m (mv)	Duration T (sec)	Bandwidth (Hz)
অ	126.30	0.50	193
আ	115.90	0.47	219
ই	43.60	0.41	628
ঈ	51.70	0.46	443
উ	79.80	0.46	341
ঊ	85.20	0.51	267
ঋ	66.80	0.48	405
এ	91.10	0.45	313
ঐ	107.70	0.44	440
ও	105.90	0.49	283
ঔ	115.70	0.45	281
অং	129.10	0.44	377
অঃ	128.40	0.41	351
ক	129.70	0.50	171
খ	124.20	0.54	186
গ	119.00	0.51	172
ঘ	125.20	0.52	196
চ	120.00	0.43	173
ছ	125.50	0.50	193
জ	122.70	0.48	192
ঝ	123.70	0.52	200
ট	129.00	0.49	179
ঠ	125.70	0.50	194
ড	122.20	0.48	183
ঢ	126.00	0.50	201
ণ	110.50	0.42	400
ত	126.20	0.50	174
থ	123.20	0.53	223
দ	128.00	0.49	172
ধ	126.50	0.52	224
ন	112.00	0.42	362

letter of the alphabet	Amplitude V_m (mv)	Duration T (sec)	Bandwidth (Hz)
প	129.00	0.48	202
ফ	124.50	0.50	211
ব	121.20	0.48	190
ভ	127.50	0.51	205
ম	118.70	0.45	322
য	120.70	0.49	191
র	124.20	0.53	156
ল	123.20	0.52	166
শ	121.50	0.45	266
ষ	119.50	0.47	272
স	116.70	0.46	260
হ	126.70	0.49	253
ড়	125.70	0.50	178
ঢ়	128.50	0.52	182
ং	124.00	0.35	369

Table IV -Average amplitude, duration and bandwidth of Bangla alphabet in female voice

Letter of the alphabet	Amplitude V_m (mv)	Duration T (sec)	Bandwidth (Hz)
অ	118.80	0.48	153
আ	108.90	0.45	166
ই	39.90	0.40	520
ঈ	47.60	0.45	359
উ	70.20	0.45	271
ঊ	79.30	0.49	191
ঋ	53.10	0.46	367
এ	86.20	0.44	238
ঐ	103.20	0.45	376
ও	102.40	0.46	188
ঔ	110.60	0.44	212

Letter of the alphabet	Amplitude V_m (mv)	Duration T (sec)	Bandwidth (Hz)
অং	117.80	0.42	309
অঃ	117.20	0.40	282
ক	124.50	0.48	137
খ	118.80	0.52	141
গ	116.20	0.49	115
ঘ	122.20	0.50	127
চ	116.50	0.41	130
ছ	117.70	0.48	132
জ	117.00	0.47	113
ঝ	119.70	0.49	146
ট	121.90	0.46	112
ঠ	121.10	0.47	129
ড	117.60	0.45	134
ঢ	119.70	0.48	139
ণ	106.30	0.40	257
ত	119.30	0.47	127
থ	117.90	0.50	140
দ	122.40	0.47	120
ধ	120.70	0.50	127
ন	107.20	0.40	231
প	119.60	0.46	114
ফ	115.70	0.48	133
ব	115.50	0.46	120
ভ	117.00	0.46	151
ম	113.90	0.42	238
য	118.30	0.46	149
র	122.20	0.51	100
ল	119.20	0.50	123
শ	117.00	0.43	144
ষ	113.30	0.43	142
স	111.40	0.42	156
হ	118.50	0.44	145
ড়	121.30	0.48	108
ঢ়	121.80	0.50	111
ৎ	116.60	0.30	285

Appendix B

Particulars of speakers

Speakers	Age (yrs)	Occupation	Talking to
M-1	28	Doctor	Female
M-2	26	Student	Male
M-3	27	Engineer	Female
M-4	20	Student	Male
M-5	30	Teacher	Male
M-6	32	Engineer	Male
M-7	23	Student	Male
M-8	40	Officer	Male
M-9	45	Lawyer	Male
M-10	29	Teacher	Male
F-1	18	Student	Male
F-2	22	Student	Female
F-3	25	Housewife	Female
F-4	12	Student	Female
F-5	38	Teacher	Female
F-6	50	Housewife	Male
F-7	20	Student	Female
F-8	20	Student	Female
F-9	21	Housewife	Female
F-10	55	Teacher	Male

** M represents male and F represents female speakers.

Appendix C

Printout of the program 'SAIF.PAS'

A program named 'SAIF.PAS' has been used to read data from the RS-232 C serial port of a PC . Computer acquisition of data has been done at an interval of 5 ms. The data thus collected has been stored in an array and finally transferred to an assigned file in a floppy diskette. The program is prepared in PASCAL. A printout of the program is presented below.

```
program read_port;
var
   aa : array [1..32767] of char;
   a,c : byte;
   count, i : integer;
   time1, time2 : integer;
   str  : string [80];
   fil : text;
   t1,t2, tsec : real;
begin
   a := port [$3fb];
   c := a and $7f;
   port [$3fb] := c;
   port [$3f9] := $0;
   port [$3fc] := $1;
   time1 := memw [$40:$6c];
   count := 0;
   repeat
       a := port [$3fe];
       count := count + 1;
       a := a and $20;
       a := a shr 5;
       a := a or $30;
       aa[count] := chr(a);
       delay (5);
       until keypressed or ( count >= 32767 );
   time2 := memw [ $40:$6c];
```

```
t1 := time1; if t1 < 0 then t1 := 65536.0 + t1;
t2 := time2; if t2 < 0 then t2 := 65536.0 + t2;
if t2 < t1 then t2 := t2 + 65536.0;
tsec := (t2 - t1) / 18.204;
tsec:=tsec/count;
write ('Enter the name of the data file '); readln (str);
assign (fil,str);  rewrite (fil);
for i := 1 to count do begin
    write (fil,aa[i]);
    if i mod 70 = 0 then writeln (fil);
    end;
writeln (fil); writeln (fil);
writeln (fil, 'Time per sample in second = ', tsec:10:6);
close (fil);
```

Appendix D

D.1 The method of least squares to fit a polynomial of degree m

If the curve is of degree m representing the polynomial

$$f = a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m \qquad \dots \qquad \dots \text{(d.1)}$$

then,

$$\epsilon_K = f_K - (a_0 + a_1 t_K + a_2 t_K^2 + \dots + a_m t_K^m) \qquad \dots \qquad \dots \text{(d.2)}$$

and $$\underline{E} = \sum_{K=1}^{n} \epsilon_K^2 \qquad \dots \qquad \dots \text{(d.3)}$$

where,

f is a function of t, m is the degree of the polynomial,
n is the number of data points,
and a_0, a_1, a_2, ... a_m are the arbitrary constants.

f is the functional value (ordinate) from equation (d.1) corresponding to t_k (the experimental t-axis value),
and f_k represents the corresponding ordinate from the experimental data.

ϵ_K=Error and

$\underline{E}$ =Summation of squares of the errors.

The application of the least square error condition results in a matrix equation of the form

$$[a] = [A]^{-1} [b] \qquad \dots \qquad \dots \text{(d.4)}$$

where,

$$[a] = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_m \end{bmatrix}$$

$$[A] = \begin{bmatrix} n & \sum_{K=1}^{n} t_K & \sum_{K=1}^{n} t_K^2 & \cdots & \sum_{K=1}^{n} t_K^m \\ \sum_{K=1}^{n} t_K & \sum_{K=1}^{n} t_K^2 & \sum_{K=1}^{n} t_K^3 & \cdots & \sum_{K=1}^{n} t_K^{m+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{K=1}^{n} t_K^m & \sum_{K=1}^{n} t_K^{m+1} & \sum_{K=1}^{n} t_K^{m+2} & \cdots & \sum_{K=1}^{n} t_K^{2m} \end{bmatrix}$$

and

$$[b] = \begin{bmatrix} \sum_{K=1}^{n} f_K \\ \sum_{K=1}^{n} f_K \cdot t_K \\ \vdots \\ \sum_{K=1}^{n} f_K \cdot t_K^m \end{bmatrix}$$

Solving equation (d.4) gives the arbitrary constants a_0, a_1, a_2, ... a_m resulting in a polynomial of the form represented by equation (d.1).

D.2 Printout of the program 'CFIT.FOR'

The program 'CFIT.FOR' is used to fit a polynomial of upto 9_{th} degree by using the method of least squares. The program is prepared in FORTRAN 77. A printout of the program is presented below.

```
C-------------------------------------------------------------------C
C     This program fits a set of data using the method of least squares.C
C-------------------------------------------------------------------C
C     It can fit a polynomial upto 9th (10 terms) degree            C
C-------------------------------------------------------------------C
      PARAMETER(ND=5,ND1=6,NDATA1=200)
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION X(NDATA1),Y(NDATA1),A(ND,ND1),D(ND),C(ND)
      OPEN(UNIT=5,FILE='INP')
      OPEN(UNIT=6,FILE='RESULT',STATUS='NEW')
      OPEN(UNIT=7,FILE='SO',STATUS='NEW')
      WRITE(*,*)'*  This program fits a set of data using the method'
      WRITE(*,*)'   of least squares. It can fit a polynomial of up-'
      WRITE(*,*)'   to ninth degree (10 terms).'
      WRITE(*,*)
      WRITE(*,*)'** If you want to fit a polynomial of degree K,then'
      WRITE(*,*)'   set ND=K+1,ND1=ND+1 in the original program '
      WRITE(*,*)'   The existing set value is'
      WRITE(*,*)'   ND=',ND
      WRITE(*,*)'***If you have,nt set the proper parameter values set'
      WRITE(*,*)'   them in the original program, make the execution'
      WRITE(*,*)'   file using (for1 filename; pas2; link filename;)'
      WRITE(*,*)'   and then run the  execution file : filename.exe '
      WRITE(*,*)
      WRITE(*,*)'# Enter number of pairs of input data(ndata)'
      WRITE(*,*)'  (No. of data must be between 2-200)'
      READ(*,*) NDATA
      WRITE(*,*)'# If you want to fit a polynomial of degree K, then'
      WRITE(*,*)'  your choice will be K+1=N ;# Maximum limit of K=9'
      WRITE(*,*)'# Please enter your choice'
      READ(*,*) N
      WRITE(*,*)'1=INPUT DATA IN FILE   2=DIRECT KEY BOARD ENTRY'
      READ(*,*) NFILE
      IF(NFILE .EQ. 1) GOTO 680
      WRITE(*,*)
      WRITE(*,*)'# Enter pairs of input data X ,Y '
```

```
      WRITE(*,*)
      DO 400 I=1,NDATA
      WRITE(*,*)'# Data (X,Y)',I
      READ(*,*) X(I),Y(I)
400   CONTINUE
      GOTO 670
680   DO 660 I=1,NDATA
      READ(5,*) X(I),Y(I)
660   CONTINUE
670   XXX=X(NDATA)
      WRITE(*,*) XXX
      WRITE(*,*)'# Enter no. of pairs of data calculated using function'
      READ(*,*) NOUT
      WRITE(*,*) NOUT
      XSTEP=(XXX-X(1))/FLOAT(NOUT)
      WRITE(*,*) XSTEP
      N1=N
      DO 440 I=1,N1
      DO 440 J=1,N1
      A(I,J)=0.0
      D(I)=0.0
440   CONTINUE
      DO 450 I=1,N1
      DO 460 J=1,N1
      IF((I+J) .GT. 2)GOTO 130
      A(I,J)=NDATA
      GOTO 460
130   DO 470 K=1,NDATA
      A(I,J)=A(I,J)+X(K)**(I+J-2)
470   CONTINUE
460   CONTINUE
      DO 480 K=1,NDATA
      D(I)=D(I)+Y(K)*X(K)**(I-1)
480   CONTINUE
450   CONTINUE
      N2=N1+1
      DO 490 I=1,N1
      A(I,N2)=D(I)
490   CONTINUE
C     Call subroutine SIMAL to solve simultaneous linear equations
      CALL SIMAL(A,C,N1,N2)
      IF(C(2) .GE. 0.0)GOTO 180
      WRITE(6,3005)C(1),C(2)
```

```
      WRITE(*,3005)C(1),C(2)
3005  FORMAT(/10X,'POLYNOMIAL FUNCTION: Y=',E12.6,E12.6,'X'/)
      GOTO 190
180   WRITE(6,3006)C(1),C(2)
      WRITE(*,3006)C(1),C(2)
3006  FORMAT(/10X,'POLYNOMIAL FUNCTION: Y=',E12.6,'+',E12.6,'X'/)
190   IF(N .EQ. 2)GOTO 210
      DO 500 I=3,N
      L=I-1
      IF(C(I) .GE. 0.0)GOTO 200
      WRITE(6,3007)C(I),L
      WRITE(*,3007)C(I),L
      GOTO 500
200   WRITE(6,3008)C(I),L
      WRITE(*,3008)C(I),L
500   CONTINUE
3007  FORMAT(40X,E12.6,'X**',I3)
3008  FORMAT(40X,'+',E12.6,'X**',I3)
C     Print input values of X and Y and calculated values of Y
210   WRITE(6,3009)
      WRITE(*,*)
      WRITE(*,3009)
3009  FORMAT(17X,'X',7X,'Y(ACTUAL)',3X,'Y(CALC)'/14X,
     +3(9(1H=),2X)/)
      S=0.0
      DO 530 I=1,NDATA
220   Y1=C(1)
      DO 540 J=2,N
      Y1=Y1+C(J)*X(I)**(J-1)
540   CONTINUE
      S=S+((Y(I)-Y1)/(NDATA-1))**2
      WRITE(6,3010)X(I),Y(I),Y1
      WRITE(*,3010)X(I),Y(I),Y1
530   CONTINUE
3010  FORMAT(10X,E12.6,2(E12.6,1X))
3012  FORMAT(E12.6,1X,E12.6)
      I=1
      XSTEP1=X(1)
      X(1)=XSTEP1
      XSS=XSTEP1
620   Y1=C(1)
      DO 605 J=2,N
      Y1=Y1+C(J)*X(I)**(J-1)
```

```
605   CONTINUE
      WRITE(7,3012)X(I),Y1
      IF(XSS .GE. XXX)GOTO 650
      I=I+1
      XSS=XSS+XSTEP
      X(I)=XSS
      GOTO 620
650   S=DSQRT(S/NDATA)
      WRITE(*,*)
      WRITE(6,3011)S
      WRITE(*,3011)S
3011  FORMAT(/10X,'# Standard deviation=',E10.5/)
      WRITE(*,*)
      WRITE(*,*)
      STOP
      END
C     Subroutine for solving simultaneous equations
      SUBROUTINE SIMAL(A,C,M,N)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(M,N),C(M)
      DO 411 K=1,M
      PIVOT=A(K,K)
      IF(DABS(PIVOT) .LE. D-10)GOTO 20
      DO 412 J=K,N
      A(K,J)=A(K,J)/PIVOT
412   CONTINUE
      DO 413 I=1,M
      IF(I .EQ. K) GOTO 413
      PIVI=A(I,K)
      DO 414 J=K,N
      A(I,J)=A(I,J)-PIVI*A(K,J)
414   CONTINUE
413   CONTINUE
411   CONTINUE
      DO 415 J=1,M
      C(J)=A(J,N)
415   CONTINUE
      RETURN
20    WRITE(6,3000)K
3000  FORMAT(10X,'Pivot element',I2,'is close to zero'/)
      RETURN
      END
```

Appendix E

On-off characteristics of English speech

The on-off characteristics of English speech are presented below by two figures and a table. They have been taken from reference [2].

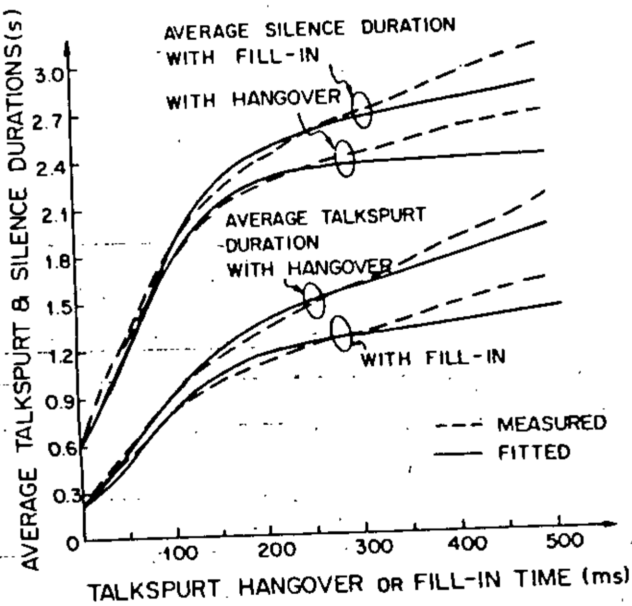


Fig.E.1 Measured and fitted average talkspurt and silence durations for different talkspurt hangover and fill-in time.

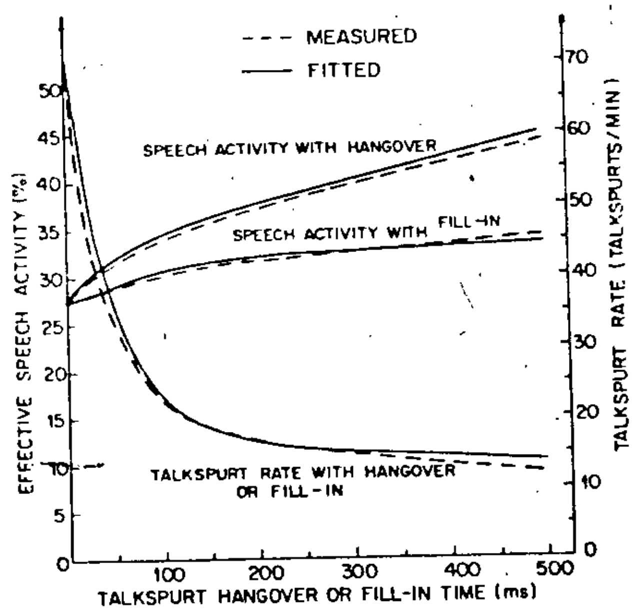


Fig.E.2 Measured and fitted speech activity and talkspurt rate for different values of talkspurt hangover or fill-in time.

Table E.1

TEMPORAL PARAMETERS OF CONVERSATIONAL ENGLISH SPEECH MEASURED WITH THE SAME LENGTH OF HANGOVER AND FILL-IN TIME (200 ms)

Speech Temporal Parameters	With hangover	With fill-in
Mean silence duration	2220 ms	2400 ms
Mean talkspurt duration	1300 ms	1110 ms
Talkspurt rate	16.67/min	16.67/min
Speech activity	37%	31.5%

References

[1] S. Z. Huque, "Statistical Analysis of Messages and their coding of Bengali Language", M. Sc. Engg. thesis, 1985, BUET.

[2] H. H. Lee, and C. K. Un, "A Study of On-Off Characteristics of conversational Speech", IEEE Trans. Commun., Vol. COM 34, June, 1986, pp. 630-636.

[3] M. A. K. Pramanik, "Acoustical study on the Vowel Structure of Bengali Language", Doctoral thesis, Tohoku University, Sendai, 1977, Japan.

[4] P. K. Das, "On information Contents of Bengali Language and Noise in Microwave Communication in Bangladesh", M. Sc. Engg. thesis, 1976, BUET.

[5] J. Bellamy, "Digital Telephony", John Wiley & Sons, New York, 1982, pp. 117, 394, 403.

[6] J. M. Fraser, D. B. Bullock, and N. G. Long, "Over-All Characteristics of a TASI system", Bell Syst. Tech. J., July, 1962, pp. 1439-1473.

[7] P. T. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation", Bell syst. Tech. J., Vol. 48, September, 1969, pp. 2445-2472.

[8] P. T. Brady, "A Technique fo Investigating On-Off Patterns of Speech", Bell Syst. Tech. J., Vol. 44, January, 1965, pp. 1-22.

[9] J. Millman and C.C. Halkias, "Integrated Electronics", International Student Edition, McGraw-Hill Book Company, Tokyo, 1986, pp. 556-557.

[10] S. Gottfried Byron, "Theory and Problems of Programming with PASCAL", McGraw-Hill Book Company, Newyork, 1988.

[11] "IBM Personal Computer XT Hardware Reference Library", pp. 1-185 to 1-213.

[12] R. G. Irvine, "Operational Amplifier Characteristics and Applications", Printice-Hall, Inc., New Jersy, 1981, pp. 69-73, 174-176.