

Chapter 3

Speech Materials and Instrumentation

To a certain extent recording of speech samples and analysis of exemplars depends on the quality and contents of the material prepared for experiments. Therefore, preparing speech materials seems an important part of the research methodology. For this particular case study, various texts have been prepared in Standard Hindi using devnagri script, including the words or phrases commonly found in telephonic interception of criminal acts. Later, these prepared texts were transliterated into each dialect chosen for the study considering the accent as well as the uses.

3.1 Selection of Informants

In order to minimize the intra-dialectal variation due to regions, informants have been selected from a uniform area of the regional dialect. Also they are chosen from the closed age group of 20–25 years with minimum higher secondary education. Condition is that the informant should not have any influence of other native language on his or her dialect. Considering the criteria mentioned above, 15 male and 15 female informants were selected from each dialect. Information related to the informants and their dialectal backgrounds are also noted. Out of the various districts in which a dialect is being spoken, only one place has been chosen to collect the samples of the informants. Name of the regional dialect and the place from where the speech samples of the informants have been collected are mentioned below in Table 3.1.

Table 3.1 Name of the dialect and the place of recording

Name of the regional dialect	Place of recording
Bundeli	Jhansi (Uttar Pradesh)
Khariboli	Anwalkhera (Uttar Pradesh)
Marwari	Alwar (Rajasthan)
Bhojpuri	Deoria (Gorakhpur)
Chattisgarhi	Charoda (Raipur)
Kanauji	Kayanpur (Kanpur)
Haryanvi	Hisar (Haryana)

3.2 Recording of Speech Exemplars

The choice of the microphone depends on the goal of a particular purpose. In this study, the microphone used is a dynamic microphone of Philips (Model DM295) with 1.8 mV/Pa sensitivity, 600 ohms impedance, and frequency range of 100–12,000 Hz.

With the microphone of above-mentioned specifications, speech samples have been recorded directly on computer using inbuilt multimedia card. Speech samples of informants (dialect speakers) are recorded in three repetitions, in Khariboli as well as in their regional dialect.

3.3 Digitization of Speech Samples

Currently, for speech sample analysis, we use computer-based equipment and data is in digital domain. All analog signals need to be converted into digital signals, where Analog-to-Digital converter (abbreviated ADC) is used as an electronic circuit that converts continuous analog signal into discrete digital numbers. Here, various types of ADCs are discussed:

A *successive approximation ADC* uses a comparator and very useful to reject ranges of voltages, eventually settling on a final voltage range. Successive approximation works constantly by comparing the input voltage with a known reference voltage until the best approximation is achieved. In entire process, a binary value of the approximation is stored in a successive approximation register (SAR) and the SAR used as a reference voltage for comparisons. Though this type of ADCs has a good resolution and quite wide range, due to more complex design they are not very popular. A *delta-encoded ADC* has an up-down counter that feeds a digital signal to analog converter (DAC). Both the input signal and the output signal of DAC go to a comparator, which controls the counter. The circuit uses negative feedback from the comparator to adjust the counter until the DAC's output is close enough to the input signal. The number is read from the counter. The great advantage with delta converters is the wide range and high resolution, except the conversion time is dependent on the input signal level.

A *ramp-compare ADC* (also called integrating, dual-slope or multislope ADC) produces a saw-tooth signal in which voltage ramps up and then quickly falls to zero. When the ramp starts, a timer starts counting. When the ramp voltage matches the input, the timer's value is recorded. Timed ramp converters require a least number of transistors. The ramp time is sensitive to temperature because the circuit generating the ramp is often just some simple oscillator. A special advantage of the ramp-compare system is that comparing a second signal just requires another comparator, and another registers to store the voltage value.

A *pipeline ADC* (also called subranging quantizer) uses two or more steps of subranging. In the first step, conversion is done and in the second step the difference to the input signal is determined with a digital-to-analog converter (DAC). This difference is then converted and the results are combined in the last step. This type of ADC is fast, has a high resolution, and only requires a small size.

A *sigma-delta ADC* (also known as a delta-sigma ADC) filters the desired signal band by oversampling of the desired signal by a large factor. ADCs are integral to much current music reproduction technology, since much music production is done on computers; even when analog recording is used, an ADC is still needed to create the PCM (Pulse Code Modulation) data stream that goes onto a compact disk. ADCs are used virtually everywhere where an analog signal has to be processed, stored, or transported in digital form.

3.4 Sampling of Speech Exemplars

The recorded utterances of the informants chosen for the study have been subjected to preliminary auditory analysis for selection of appropriate speech data from the raw data. The utterances are chosen in which the accent features of the speakers are well reflected and are found suitable on the basis of speech quality/clarity of the recorded sample. Speech exemplars of 15 male informants and 15 female informants are chosen in each dialectal group.

3.5 Instrumentation

3.5.1 Microphones

A microphone is a transducer that converts sound energy into electrical energy. Sound information exists as patterns of air pressure and the microphone changes this information into patterns of electric current.

There are a variety of mechanical techniques that can be used to build a microphone. The two most commonly used methods are the magnetodynamic and variable condenser. A majority of microphones used in recording of sounds are either capacitor (electrostatic) or dynamic (electromagnetic) models, which employ a

moving diaphragm to capture the sound, but make use of a different electrical principle for converting the mechanical energy into an electrical signal. The efficiency of this conversion is very important because the amount of acoustic energy produced by voices and musical instruments is very small.

3.5.1.1 Dynamic Microphone

In the magnetodynamic, commonly called dynamic microphone, sound waves cause movement of a thin metallic diaphragm and an attached coil of wire. A magnet produces a magnetic field, which surrounds the coil and motion of the coil within this field causes the current to flow. It is important to remember that the current is produced by motion of diaphragm and the amount of current is determined by the speed of that motion. The problem with dynamic microphones is that they are most effective only while working with relatively loud sound sources.

3.5.1.2 Ribbon Microphone

These microphones are comprised of a thin metal ribbon suspended in a magnetic field. When the sound energy is encountered, the electrical signal generated is induced in the ribbon itself. The main advantage of ribbon microphone is its smooth and detailed sound, but this type of microphones is costly and more fragile than conventional dynamic microphones.

3.5.1.3 Capacitor Microphone (Condenser Microphone)

In a condenser microphone, the diaphragm is mounted close to backplate without touching it. The voltage of the battery, the area of the diaphragm, and the backplate and the distance between the two determine the amount of charge. When the distance changes in response to the sound, the current flows in the wire as the battery maintains the correct charge. The amount of the current is essentially proportional to the displacement of the diaphragm and is so small that it must be electrically amplified before it leaves the microphone.

The wide availability of electret condenser microphones has greatly simplified the problem of obtaining high-quality recordings. Electret microphones respond directly to the sound pressure of the speech signal. Directional electret microphones respond differentially to sounds coming from one direction. This can be an advantage if one is recording the samples in a noisy environment.

3.5.2 *Sound Spectrograph*

Alexander Melville Bell in 1867 developed a visual representation of the spoken words later named as “visible speech” at Bell Telephone Laboratory. In 1940s Potter,

Kopp, and Green developed a new method of speech sound analysis using speech spectrograph. Dr. Ralph Potter introduced an electromechanically acoustic spectrograph in 1941. In 1962 Lawrence Kersta, an engineer and a staff member of the Bell laboratories, reexamined the voiceprint method at the request of law enforcement group and introduced the instrument named sound spectrograph as a potential tool for Forensic Speaker Identification. Basic function of this device was to convert the speech into visual representation of its frequency and intensity components.

A sound spectrograph has four parts: a magnetic recorder, electronic filters, a rotating drum on which the spectrogram is recorded, and an electrically operated stylus. The traditional analog version of the sound spectrograph records the input signal on a magnetic medium that goes round the outside edge of the thin drum. Magnetic image is formed on the thin recording disk by the recording head just like a conventional tape recorder when the sound spectrograph switch is on recording mode.

The voice/sound spectrograph is of three types:

Analog spectrograph

Digital spectrograph

Hybrid spectrograph

In ***analog spectrograph*** speech from the microphone is fed into a band pass filter. Harmonics of the voice whose frequency falls within the range of that filter gives the output with the amplitude proportional to its strength, and then produces a three-dimensional record with the stylus on a paper, showing the change in frequency and amplitude with time.

A ***digital spectrograph*** consists of special circuits embedded in the microprocessor systems to produce the spectrogram simultaneously with the speech. Voice identification Inc., USA, has produced a real-time digital spectrograph, which produces video display of spectrograms. The spectrograms determine the duration of the speech segments, calculate fundamental frequency and formant ranges, etc.

Hybrid spectrograph is the combination of the two spectrograms mentioned earlier.

The sound spectrograph analysis requires only a limited amount of utterance at a time, which could be recorded in one revolution of the drum. It converts the speech signal into a visual spectrum in the form of traces of the graph. The dimensions and the intensity of the traces are dependent on the utterance being analyzed. Nowadays, the use of computers in conjunction with the spectrograph has increased the volume of the recording.

3.5.3 Computerized Speech Laboratory

Computerized Speech Lab (CSL) for Windows is a hardware and software system for the acquisition, acoustic analysis, display, and playback of speech signals. It records, edits, and quickly analyzes the speech signal. It is possible to carry out the detailed studies of the utterances through segmentation of the recordings.



Fig. 3.1 Computerized Speech Laboratory Model 4300B

CSL is suitable for any acoustic signal characterized by changing spectra over time. It is a Windows-based program, which requires a computer operating under Windows 95 and Windows 98. Operations of CSL include acquisition, storing speech to disk memory, graphical and numerical display of speech parameters, audio output, and signal editing. A variety of analyses, namely, spectrographic analysis, pitch contour analysis, LPC analysis, Cestrum analysis, FFT and Energy contour analysis, etc., can be performed through this instrument. It gives results easily and quickly in comparison with the old speech spectrograph and can handle large speech data at a time. Speech exemplars chosen during the sampling process have been analyzed using Computerized Speech Laboratory model 4300 B as shown in Fig. 3.1.