An Introduction to Speech Sciences (Acoustic Analysis of Speech)

Rezaei Nasser , BSc, P.G.D., MA, PhD

Assistant Professor; University of Social Welfare and Rehabilitation Sciences,
Salehi Abolfazl , BSc , MSc
Lecturer , University of Social Welfare and Rehabilitation Sciences, Tehran.Iran

Abstract

Speech sciences deal with the acoustical characteristics of speech by means of sophisticated soft wares as well as hard wares. Although, a speech science is a well known science in the developed countries, especially the western societies, however, it has been remained almost unknown in Iran, though, in recent years a group of scholars have been involved in this branch of science. The application of the speech sciences in the field of 'Speech and Language Therapy' has been widely accepted throughout the world. The main focus of these applications is the therapeutic aspects of the program, which is used for both adults and children. However, the non-therapeutic applications are mainly used for research purposes only.

In this article, a brief introduction about the nature as well as the application of speech analysis for both normal and abnormal speeches is given. In addition, examples of the acoustical characteristics fconsonants and vowels such as, voice onset time (VOT), formant frequencies and spectrograms will be presented.

Key words:
Acoustic analysis, VOT, Spectrogram, Formant frequency, Disordered speech

Introduction

Speech science is the study of all the factors involved in producing, transmitting, perceiving and comprehending speech, including all relevant aspects of anatomy, physiology, neurology and acoustics, as well as phonetics. Speech analysis began in 1940s (i.e., during the Second World War) in the United States of America. The study of speech production from an acoustical point of view provides the means for looking at a very complex process in a simpler way. Speech science is the study of all the factors involved in producing, transmitting, perceiving and comprehending speech, including all relevant aspects of anatomy, physiology, neurology and acoustics, as well as phonetics.

Acoustics is the branch of physics that deals with sound, and Psychoacoustics is the study of the psychological response to sound; it is a division of psychophysics, or the general study of psychological responses to physical stimuli. The study of speech acoustics has both a physical and psychological side. The physical side pertains to the physical structure of the sounds of speech. The psychophysical side is concerned with the perception of these sounds. A proper understanding of speech requires knowledge of both of these aspects of speech acoustics. A well known riddle asks, "If a tree falls in the forest, but there is no one to hear it, does it make a sound?" Of course, the answer depends on the definition of terms. If sound is defined with respect to human perception, then no sound could be verified. But if sound is defined as a physical disturbance in the air, then sound must have occurred. A riddle more regarding speech might be: "If speech is made visible as patterns on paper (or a video monitor), but no one hears it, is it really speech?" Sound is vibration. Vibration is a repetitive to-and-fro motion of a body. Usually we do not directly hear the vibration in a sound source such as an engine, but rather we hear the vibrations that are broadcast, or transmitted, in a medium like air. Figure B-1 shows a simple physical arrangement to demonstrate the nature of sound. The sound source is a stretched rubber band that can be pulled to set it into vibration. The band undergoes a series of to-and-fro vibration after it is pulled. The initial vibrations are of large amplitude, meaning that the to-and-fro swings have a relatively great motion. The amplitude diminishes until motion eventually stops altogether. The reduction in amplitude reflects damping, or the loss of energy.

Oscillograms:

The first major advance in the acoustic analysis of speech began with Oscillograms (waveforms, or graphs of amplitude over time) which was an important advance. In this method, the sounds selected for analysis were often vowels, because they are relatively easier to analyse than most consonants.

The Henrici Analyser:

Is one of the earliest tools for spectral analysis. It is a mechanical device consisting of five rolling integrating Units (glass spheres). They could obtain Oscillograms, tracing on a surface, calculating the values of amplitude, and pressure.

Filtering:

A filter is a frequency-selective transmission system. That is, a filter will pass energy at certain frequencies but not at others. A filter allows a selective look at the energy in various frequency regions.

The Spectrograph:

Provided major advantages to the study of speech and was developed in 1940s. It was fast, provided a better description of the energy concentration in speech, and finally, it produced a running short-term spectrum, enabling the scientists to visualise how energy concentrations changed in time. The display of the running short-term spectrum is called spectrogram.

Digital Signal Processing of Speech:

The dominance of the spectrograph was seriously challenged only with the introduction of digital computers. The challenge has intensified with the continued refinement of computers (hardware) and analysis programs (software). Some developments in the use of digital methods for speech analysis are: waveform display; waveform editing; playback of original or edited signals; F0; Jitter or Shimmer analysis; spectrogram; FFT' LPC; and speech synthesis.

The Speech Chain:

Understanding of subject of spoken communication requires knowledge of many disciplines. The Speech Chain provides an easy-to-understand introduction to these topics and explains how events within these domains interact to bring about communication by speech. A convenient way of examination what happens during speech is to take the simple situation of two people talking to each other. For example, you as a speaker want to transmit information to another person, the listener. The first think you have to do is to arrange your thoughts, decide what you want to say and then put what you want to say into linguistic form. The message is put into linguistic form by selecting the right words and phrases to express its message, and by placing these words in the order required by the grammatical rules of the language. This process is associated with activity in the speaker's brain, and it is from the brain that appropriate instructions, in the form of impulses along the motor nerves, are sent to the muscles that activate the vocal organs (the lungs, the vocal cords, the tongue, and the lips). The nerve impulses

set the vocal muscles into movement, which in turn, produce precise pressure change in the surrounding air. We call this pressure change a sound wave. Sound waves are often called acoustic waves, because acoustics is the branch of physics concerned with sound.
The movement of the vocal organs generate a speech sound wave that travels the air between speaker and listener. Pressure changes at the ear activate the listener's hearing mechanism and produce nerve impulses that travel along the acoustic nerve to the listener's brain. In the listener's brain, a considerable amount of activity is already taking place, and this activity is modified by the nerve impulses arriving from the ear. This modification of brain activity, in ways that are yet fully understood, brings about recognition of the speaker's message. We see, therefore, that speech communication consists of a chain of events linking the speaker's brain with the listener's brain. We shall call this chain of events the speech chain.

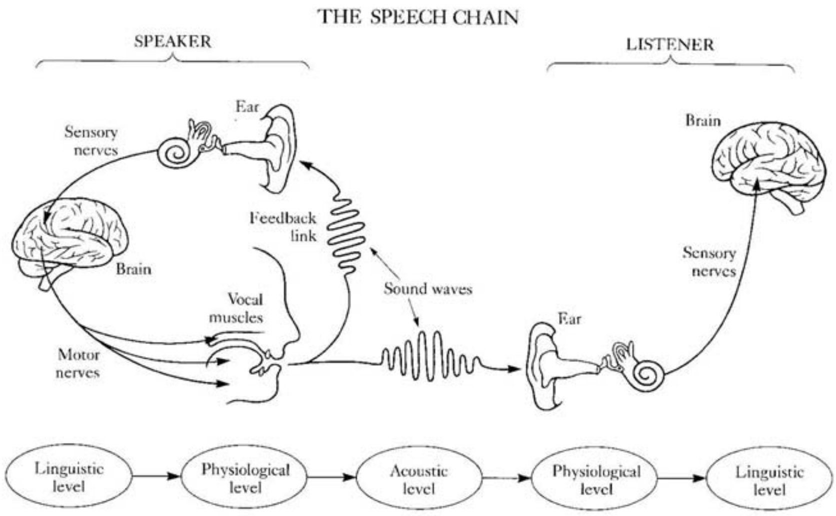


FIGURE 1.1 The speech chain: the different forms of a spoken message in its progress from the brain of the speaker to the brain of the listener.

Acoustic Analysis of Speech

Acoustic analysis of speech is the study of the acoustical characteristics of speech, both normal and abnormal speech. It involves the physical aspects of spoken language. These include; waveform analysis, FFT or LPC analysis, voice onset time (VOT) measurements, formant frequency measurements, and so on.

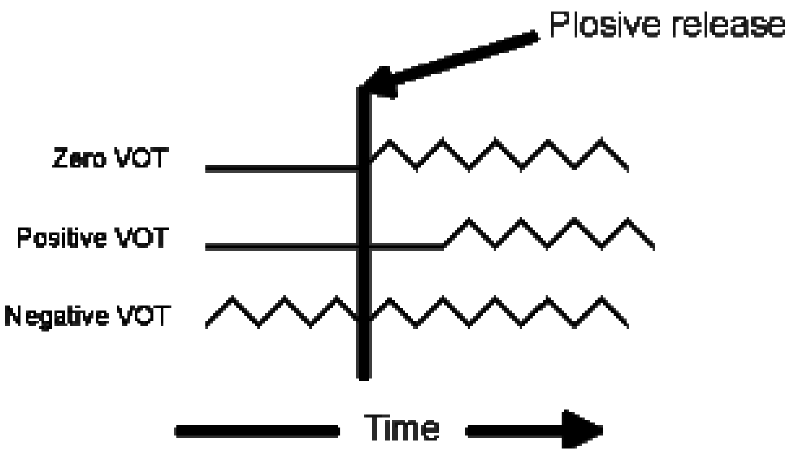
Some of these properties will be reviewed briefly.

Voice Onset Time (VOT):
Voice Onset Time (VOT) for instance, is the duration of the period of time between the release of a plosive and the beginning of vocal fold vibration. This period is usually measured in milliseconds (ms), which is a microscopic analysis. It is useful to distinguish at least three types of VOT which are shown in the schematic diagram below:

-Zero VOT: where the onset of vocal fold vibration coincides (approximately) with the plosive release.

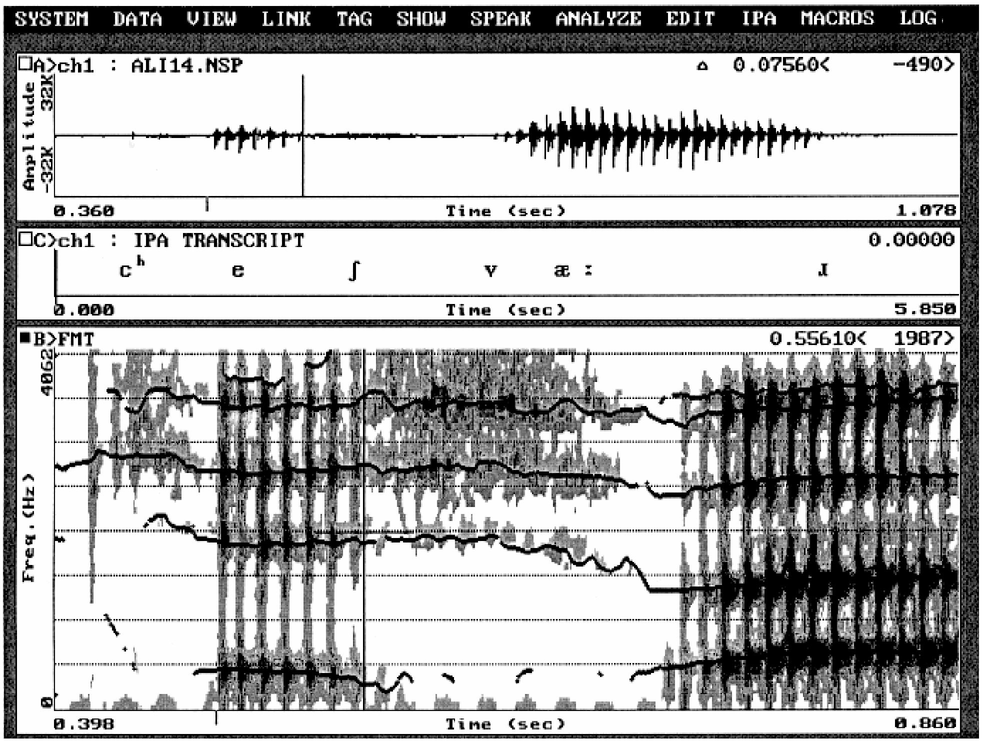
-Positive VOT: where there is a delay in the onset of vocal fold vibration after the plosive release.

-Negative VOT: where the onset of vocal fold vibration precedes the plosive release.



Formant frequency

Formant is the vocal tract resonance. Theoretically, there is an infinite number of formants, but for practical purposes only the lowest three or four are of interest.
Formants are identified by formant numbers (e.g., F1, F2, F3, and F4, numbered in succession beginning with the lowest frequency formants). Each formant can be described by two characteristics: centre frequency (commonly called formant frequency) and bandwidth (formant bandwidth, which is a measure of the width of energy in the frequency domain.



Acoustical Characteristics of Vowels

Producing different vowels is like altering the size and the shape of the vocal tract, which is like a tube. The air in this tube is set in vibration by the pulse of air from the vocal folds. Every time they open and close, the air in the vocal tract above them will be set in vibration. Because, the vocal tract has a complex shape, the air within it will vibrate in more than one way. Usually we can consider the body of air behind the raised tongue (i.e., in the throat) to vibrate in one way, and the body of air in front of the tongue (i.e., in the mouth) to vibrate in another way. In the vowel /i:/ as in deed, the air behind the tongue will vibrate at about 250 Hz, and the air in front of it at about 2100 Hz.

The relationship between the tongue position and formant frequencies

The frequency of F1 is inversely related to tongue height (e.g., high vowels have a low F1 frequency), and the frequency of F2 is related to tongue advancement (e.g., F2 frequency increases as the tongue position moves forward in the mouth).

F1: High vowels such as /i/ and /u/ have a relatively low frequency of the first formant (F1), whereas the low vowels /a/ and /ae/ have a relatively high frequency of this formant, that is, the frequency of

F1 varies inversely with tongue height of the vowel.

F2: On the other hand, back vowels /u/ and /a/ share a low frequency of F2, whereas the front vowels /i/ and /ae/ have a high F2, that is, the frequency of the second formant varies with the posterior-anterior dimension of vowel articulation.

This result points to an articulatory-acoustic correspondence: the frequencies of the first two formants, F1 and F2 can be related to dimensions of vowel articulation.

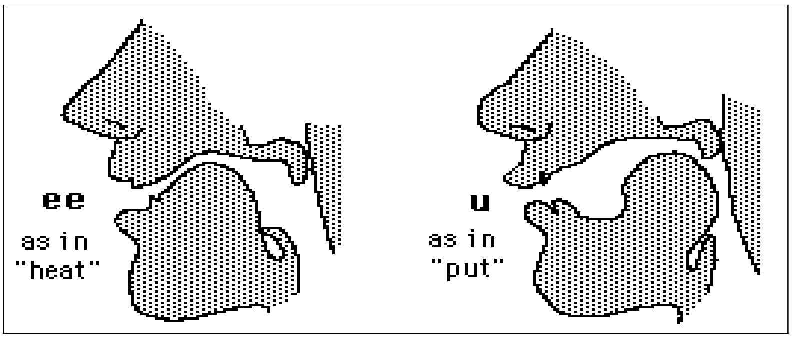


Figure 4: Illustration of the tongue position for the production of the vowels /i:/ and /u/.

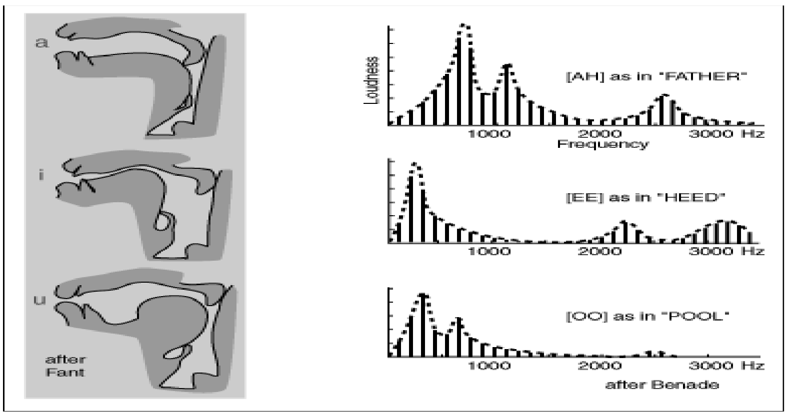


Figure 5: The relationship between the tongue position and formant frequencies for different vowels.

The Acoustic Characteristics of Disordered Speech

The Characteristics of disordered speech are somewhat different from those of the normal speech. These include:

Howell and Vause (1986), in an attempt to investigate the acoustic properties of stuttered speech, analysed the spectral properties of 8 English vowels in adult stutterers in comparison with the subsequent fluent vowels. They intended to examine some of the previous researchers findings such as Van Riper (1971) who had stated that, the stutterer produces the schwa vowel at a point where some other vowel is required. In this experiment, words consisting of a voiceless consonants in a CVC syllable were investigated, where the final consonant in the fluent word includes nasals, fricatives (voiced and voiceless),

a)Vowels & Diphthongs
b)Sonorants: nasals, glides, liquids
c)Obstruents: e.g., stops, fricatives

Dysfluent vs. Fluent Vowels and voiceless plosives

The acoustic properties and formant frequencies of stutterers for fluently and dysfluently produced vowels, as well as the duration and amplitude of vowels were examined. The results revealed that dysfluent vowels were shorter than the fluently produced vowels. Other acoustic data showed that the brief vowels in repetitions had formant locations at the same frequencies as in the fluent speech of the same speaker. Stuttered vowels were low in amplitude and short in duration, which caused stuttered vowels to be perceived as schwa.

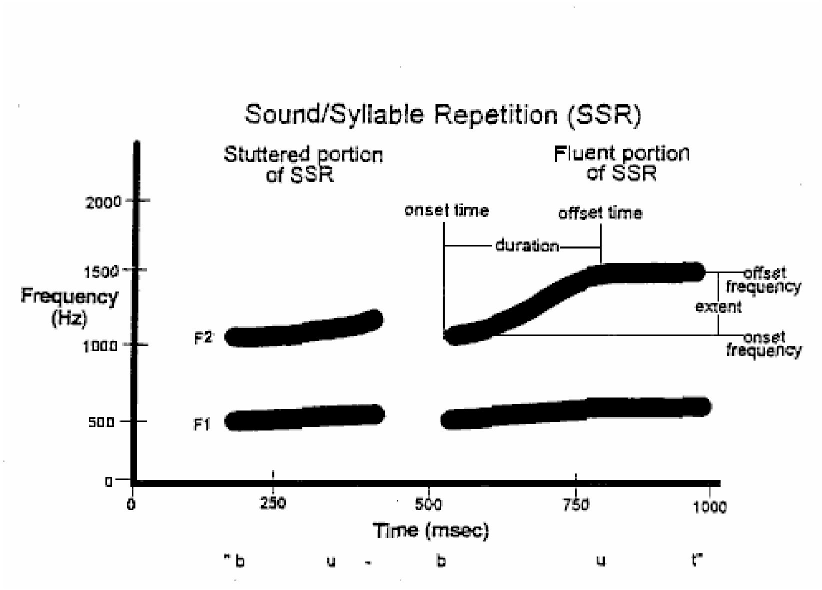


Figure 6: Idealised spectrogram of a young male who stutters producing a sound/syllable repetition (SSR) on the word "but" with one iteration of "bu-" in the stuttered or repeated portion of the SSR ("bu-but"). The stuttered portion is immediately followed by the fluent portion used for stuttered versus fluent comparisons (Yaruss and Conture, 1993, p. 888).

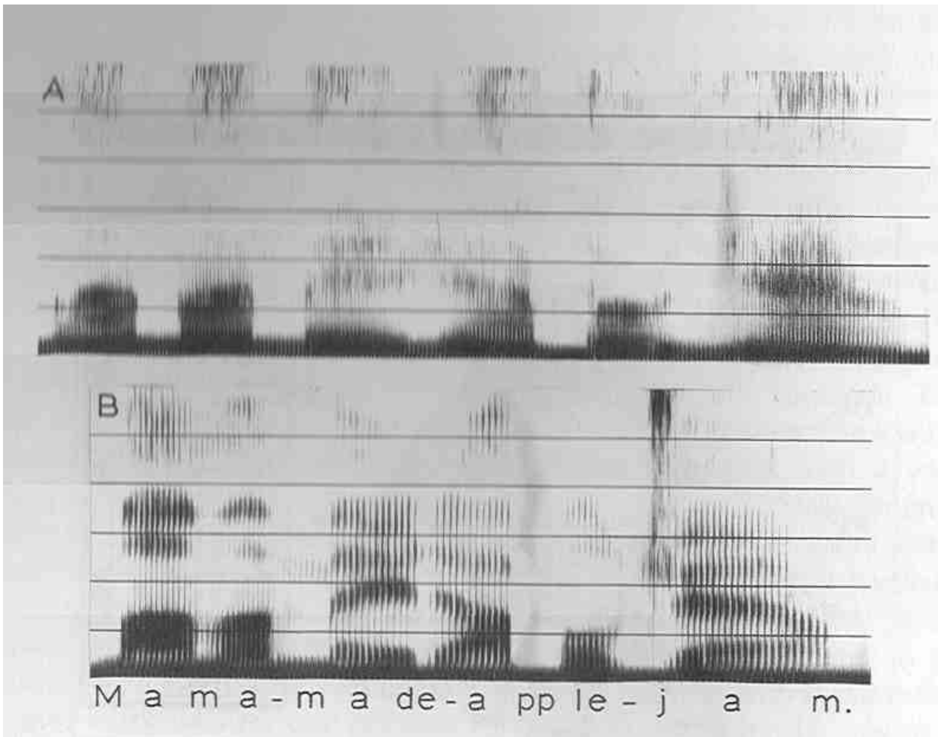


Figure 7: Spectrograms for the sentence "mama made apple jam" produced by a a speaker with velopharyngeal incompetence (hepernality) at the top (A) and by a normal speaker at the bottom (B).

References:

Ball, M. J. and Code, C. (1997). Instrumental Clinical Phonetics. England, Whurr Publishers.

Denes, P. B., and Pinson, E. N. (1993) The Speech Chain: The Physics and Biology of Spoken Language. New York, W. H. Freeman and Company.

Fry, D. B. (1982). The Physics of Speech. Cambridge: Cambridge University Press.

Howell, P., and Vause, L. (1986). Acoustic Analysis and Perception of Vowels in Stuttered Speech. Journal of the Acoustic Society of America, 79 (5), 1571-1579.

Kent, R. D., and Read, C. (1992). The Acoustic Analysis of Speech. USA, Singular Publishing Group, Inc.

Rezaei-Aghbash, N. (2003). Acoustic Characteristics of Dysfluent Speech: A Study of Persian and English stutterers. An unpublished PhD dissertation. UK.

Van Riper, C. (1971). The Nature of Stuttering. 2nd edition. NJ (USA), Prentice-Hall, Englewood Cliffs.

Yaruss, J. S., and Conture, E.G. (1993). F2 Transition During Sound/Syllable Repetitions of Children Who Stutter.