Study of Tones Characteristics in Thai Language using Fast Fourier Transform (FFT)

Janchai YINGPRAYOON
janchai.yi@ssru.ac.th or dr.janchai@gmail.com
International College,
Suan Sunandha Rajabhat University, Bangkok,
THAILAND

Abstract:
The Thai language is a tonal *language. Tones are the core of the language. Tones distinguish the meaning of one word from another. Thai language is one of the languages that uses tones for communication of information. Thai language uses consonant sound symbols to indicate how the words should be pronounced correctly. Otherwise we cannot understand the meaning of the words if they have the same tones but different symbols. In order to speak Thai correctly in terms of meaning the tones of the words have to be pronounced correctly. This work is to study the tone characteristics of Thai language using Fast Fourier Transform (FFT) to analyze the frequency patterns of the 5 tones in Thai language.*

1. Introduction

Spoken language communication is the most important activity that distinguishes humans from animals. While many animal species communicate and exchange information using sound, humans are unique in the complexity of the information that can be conveyed using speech, and in the range of ideas, thoughts and emotions that can be expressed.

Despite the importance of speech communication for the entire structure of human society, there are many aspects of the speech communication process that are not fully understood. Research on speech and language is typically carried out by different groups of scientists working on separate aspects of the underlying functional and neural systems. Research from an auditory perspective focuses on the acoustical properties of speech sounds, the representation of speech sounds in the auditory system and how that representation is used to extract phonetic information. Research from psycholinguistic perspectives studies the processes by which representations of meaning are extracted from the acoustic–phonetic sequence, and how these are linked to the construction of higher-level linguistic interpretation in terms of sentences and discourse. However, there has been relatively little interaction between speech researchers from these two groups.

Various languages for oral communication in the world have different structures and patterns. The characteristics or structures and patterns of oral language will be the major factor for effective communication of information. The important structure of many languages in the world are the tones of the language. Different tones of the words in the oral language may have different meanings.

Thai language is one of the languages that uses tones for communication of information. Thai language uses consonant sound symbols to indicate how the words should be pronounced correctly. Otherwise we cannot understand the meaning of the words if they have the same tones but different symbols. In order to speak Thai correctly in terms of meaning the tones of the words have to be pronounced correctly. This may be different from English language.

This work is to study the tone characteristics of Thai language using Fast Fourier Transform (FFT) to analyze the frequency patterns of the 5 tones in Thai language.

2. Related Work

The paper by Young [1] describes about the of speech sounds in the auditory nerve of the central nervous system, focusing especially on vowel sounds. The experimental data are derived mainly from animal models (especially the cat). In terms of the human auditory system, some caution is needed in interpreting the results. It seems probable that the early stages of auditory processing are similar across all mammals. A key feature of the representation of sounds is that it is tonotopic; speech signals are decomposed into sinusoidal frequency components or groups of components and different frequency components are represented in different populations of neurons. The spectrum of the sound is represented in the relative amount of neural activity in neurons that are tuned to different frequencies. Young argues that it is probable that the form of the representation at the auditory cortex is fundamentally different from the representation at lower levels, in that stimulus features other than the distribution of energy across frequency are analyzed.

The paper by Moore [2] reviews basic aspects of auditory processing. The frequency selectivity of the auditory system refers to the ability to resolve the sinusoidal components in complex sounds, and is closely related to the tonotopic representation described by Young. Moore describes how frequency selectivity can be quantified using masking experiments. The pitch of speech sounds is related to their fundamental frequency, which is in turn related to the rate of vibration of the vocal folds. Moore describes the mechanisms by which the auditory system extracts the pitch of speech sounds and the role that pitch patterns play in speech perception, especially the perception of intonation.

The paper by Diehl [3] considers further the robust nature of speech perception. Diehl considers how the acoustical and auditory properties of vowels and consonants help to ensure intelligibility. The properties of speech sounds can be understood by considering the sounds as resulting from a source of sound energy, such as vibration of the vocal folds or turbulence produced by forcing air through a narrow constriction, followed by a filter (the vocal tract) which modifies the spectrum of the source. Diehl points out that certain types of speech sounds (e.g. the resonance patterns or ‘formant’ frequencies of specific vowel sounds) occur commonly in the languages of the world, while others occur much more rarely.

Quantal theory Stevens [4] is based on the fact that nonlinearities exist in the mapping between articulatory (i.e. vocal-tract) configurations of talkers and acoustic outputs. Given these regions of acoustic stability and instability, quantal theory is based on the idea that preferred sound categories are selected to occupy the stable regions and to be separated by unstable regions. Dispersion theory [5], like quantal theory, is based on the idea that speech sound inventories are structured to maintain perceptual distinctiveness. A vowel or consonant inventory is said to be maximally distinctive if the sounds are maximally dispersed (i.e. separated from each other) in the available ‘phonetic space’. Diehl discusses the strengths and limitations of each theory, and proposes that certain aspects of the two theories can be unified in a principled way so as to achieve reasonably accurate predictions of the properties of preferred sound inventories.

Typically, frequencies in the range of 50Hz and upwards are generated in human speech. The majority of the energy is concentrated between 300Hz and 3k Hz. The human ear, on the other hand, can detect sounds over a range of frequencies from around 20 Hz to 20 kHz. with most sensitivity in the region between about 300Hz. and 10kHz. With the account of these factors along with functional testing the frequency range of 300Hz to 3.4kHz has been found to be the most important for speech intelligibility and speech recognition.

Reducing this (300Hz. to 3.4KHz) bandwidth can significantly reduce the speech intelligibility however increasing it has been found not to significantly improve recognition or the intelligibility. Please note that increased bandwidth will improve overall sound quality however the incremental gains in sound quality have to be weighed against increased frequency usage.

The frequency band of 300Hz to 3.4kHz is therefore used in our everyday telephone system. In reality this range of bandwidth provides exceptionally understandable speech and has been the basis of our society's telephony equipment for many decades. (Nortel Networks 2001).

3. Pronouncing the Tones in Thai Language and Linguistic Analysis

Tone refers to a change in the pitch of a syllable during its pronunciation. In Thai, every syllable is pronounced in one of five tones: **Mid, Low, Falling, High and Rising.** The tone must be spoken correctly for the intended meaning of a word to be understood. Since every word has a particular mandatory tone, we say that the Thai language has obligatory lexical tone.

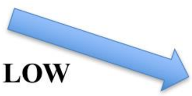
In order to analyze the frequency characteristics of these 5 tones: **Mid, Low, Falling, High and Rising**, the word “KA” will be pronounced with different tones.

Software Audio tools from Studio Six Digital, U.S.A. [6] to analyze the audio signals was used. FFT, or Fast Fourier Transform takes a time-varying input signal and transforms it into a frequency spectrum. The FFT algorithm is a mathematical procedure that breaks a signal into frequency bins. Each bin is the same size in Hertz. The size of the FFT determines the width of the bins in Hertz. The bin width = sample rate / FFT size. For example, running at the maximum sample rate of 48,000 samples/second, a 1,024 point FFT would result in each frequency bin being 48 Hz wide. This gives us great resolution at 10 kHz, but poor resolution at 32 Hz.

Software tools for analysis of Fast Four Transform called FFT Analyzer is used to analyze the voice signals. This FFT Analyzer can provide the choice of FFT size (4,096 points to 32,768 points). But in this case I used 16,384 points. This is high enough to get the good results. The voice signals of the 5 tones are recorded and analyzed by FFT. The results of FFT analyzer are shown in the figure 2. The voices of different persons will give different frequency spectrum. But the patterns of voice spectrum is the same only the variations of the frequency set. The figure 2. Shows a typical set of voice spectrum of the five tones.

MID

: spoken in a constant pitch in normal vocal range. Do not vary the pitch as the syllable is pronounced. Mid tone gives a peak frequency of about 113 Hz, equivalent to music note of A2 in the piano keyboard diagram in figure 1.



LOW

: spoken in a constant or slightly falling lower pitch, starting at a pitch lower than normal vocal range. Low tone gives a peak frequency of about 93 Hz, equivalent to music note of F#2 in the piano keyboard diagram in figure 1.

: starting slightly above comfortable speaking range, rise just a little before falling below the starting point. Falling tone gives a frequency band ranging from about of 128 Hz to 88 Hz, equivalent to music notes of C3 to F2 in the piano keyboard diagram in figure 1.

: spoken at the top of normal vocal range, producing a somewhat stressed sound. High tone gives a peak frequency of about 138 Hz, equivalent to music note of C#3 in the piano keyboard diagram in figure 1.

: rising sound, as in a question spoken in English. Rising tone gives a frequency band ranging from about of 91 Hz to 145 Hz, equivalent to music notes of F#2 to D3 in the piano keyboard diagram in figure 1.

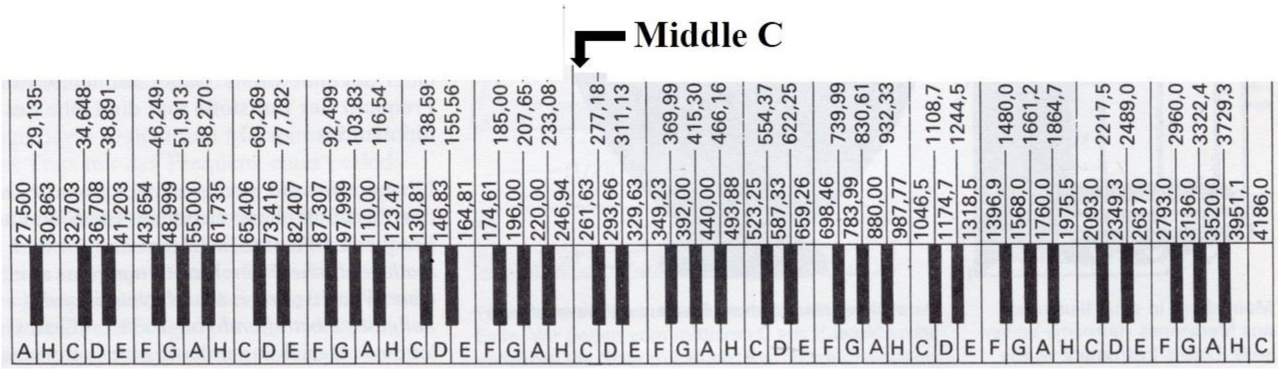


Figure 1*. Piano Keyboard Diagram [7]*

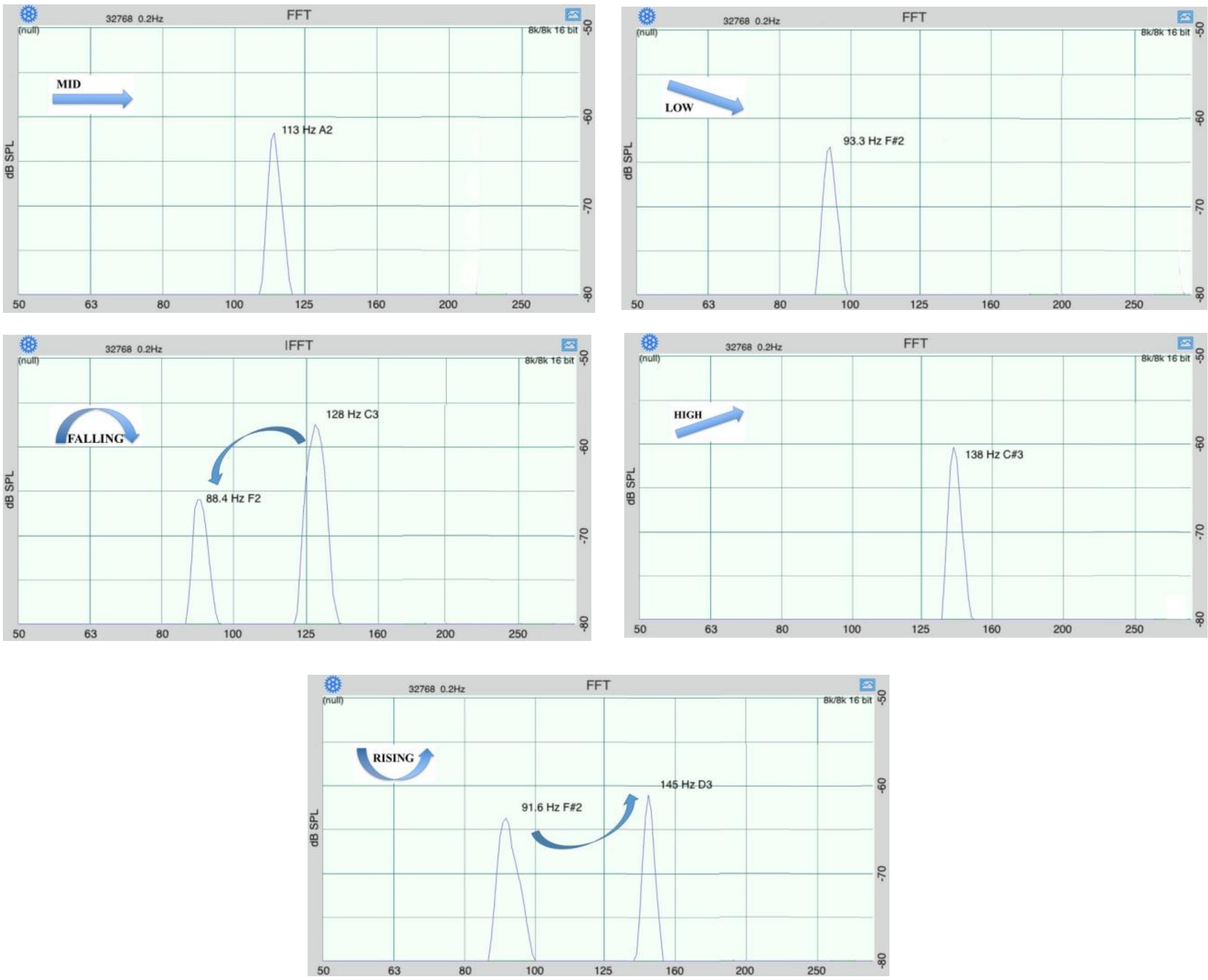


Figure 2. *FFT frequency spectrum of the 5 tones in Thai language:*

The five voice **Conclusion**

The Thai language is a tonal language. Tones are the core of the language. They are essential as important as any vowel or any consonant. Tones distinguish the meaning of one word from another.

Each syllable is pronounced with one of five distinct tones – **Mid, Low, Falling, High** and **Rising**, The middle tone starts at a middle pitch level of about 113 Hz (typically). The low tone starts low about 93 Hz. The falling tone starts high and falls to a low pitch, from 128 Hz to 88 Hz. The high tone rises at 138 Hz. The rising tone starts at low-level and gradually rises from 91 Hz to 145 Hz. Figure 3 shows a more technical analysis of the pitch (frequency) of over time.

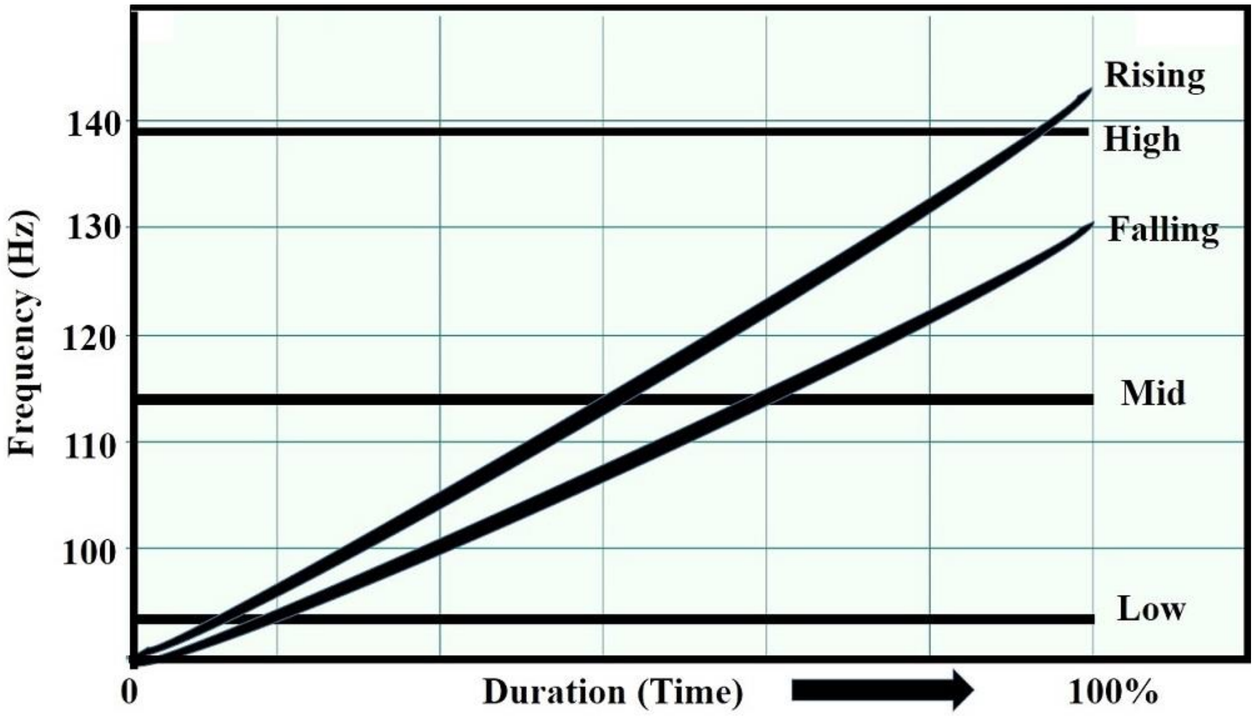


Figure 3. *Frequency diagram of tones in Thai language*

Comparing Thai and English, prosody is used for different functions. In Thai, tone has a semantic function; words which are similar but pronounced with different tones have different, unrelated meanings. English generally does not use tone in this way, but for illustration it may be useful to examine a different type of prosody in English. Consider the case of stress. In other languages like Chinese, they are also tone languages. The same study can also be done to investigate the interesting characteristics for further research.

References

[1] Young E.D. Neural representation of spectral and temporal information in speech. Phil. Trans. R. Soc. B. 2008;363:923–945. doi:10.1098/rstb/2007.2151 [PMC free article] [PubMed]

[2] Moore B.C.J. Basic auditory processes involved in the analysis of speech sounds. Phil. Trans. R. Soc. B. 2008;363:947–963. doi:10.1098/rstb/2007.2152 [PMC free article] [PubMed]

[3] Diehl R.L. Acoustic and auditory phonetics: the adaptive design of speech sound systems. Phil. Trans. R. Soc. B. 2008;363:965–978. doi:10.1098/rstb/2007.2153 [PMC free article] [PubMed]

[4] Stevens K.N. On the quantal nature of speech. J. Phon. 1989;17:3–45.

[5] Liljencrants J, Lindblom B. Numerical simulation of vowel quality systems: the role of perceptual contrast. Language. 1972;48:839–862. doi:10.2307/411991

[6] Software Audio tools from Studio Six Digital, U.S.A.

[7] Pierce J.R., KLANG: Musik mit den Ohren der Physik,Spektrum der Wissenschaft, 1983.