Gender and Situational Variation of Pitch Dynamism in Thai Speakers

Athit Wu
Faculty of Arts
Chulalongkorn University

Received: 8 Apr, 2024 *Revised: 31 May, 2024* *Accepted: 11 Jun, 2024*

Abstract

Human physiology confirms that male and female voices exhibit different fundamental frequencies (F0) intrinsically. However, whether the use of F0 in everyday speech differs between males and females remains contentious. Perceptually, female speech is often considered more fluctuating, swoopy, and melodic compared to male speech. This study aims to investigate this issue by revisiting methods of measuring F0 in utterances to determine whether the differences between male and female speakers mentioned above exist. Pitch dynamism, defined as a two-dimensional value (x, y) comprising pitch perturbation and pitch variance, has been selected as the acoustic variable. Additionally, this study examines the influence of situational context on utterance F0. Recordings were collected from six native Thai students (three males and three females), involving both reading speech and spontaneous speech. The results indicate a significant difference in utterance F0 between male and female speakers, with pitch variance serving as the primary distinguishing cue. Furthermore, the analysis reveals that situational context significantly affects the realization of utterance F0.

Keywords: F0; pitch dynamism; gender variation; situational contexts; Thai language

Introduction

Sex[1] is an important factor in the study of phonetic variation, with gender as a variable. Sound pronunciation and articulation are mainly the consequences of vocal tract anatomy. After the air is inhaled and leaves the lungs, it passes through the vocal folds, causing the membranes to vibrate. This vibration generates a wave, and the frequency of the vibration is called the fundamental frequency (also: F0), which correlates with the pitch of the voice according to the law of vibration: if an object creating a sound wave has a high vibration rate, it is associated with a high pitch. Ohala

[1] Noted that sex and gender refer to different concepts. Sex mainly refers to biological differences, while anything resulting from sex is primarily influenced by physiological factors. Gender, on the other hand, refers to the sexual identity one performs in society, regardless of social norms and expectations. Eckert and McConnell-Ginet (2003) regard sex as a social elaboration of biological differences.

(1984) proposed that voices produced by living beings are innate, having a genetic disposition, and there is a dimorphism, a condition where the two sexes of the same species exhibit different characteristics beyond the differences in their sexual organs, in the vocal tract anatomy. Subsequently, Titze (1989) drew a comparison between male and female larynxes in size, vocal fold membranous length, elasticity of tissue, and pre-phonatory glottal shape. They cited Kahane's (1978) findings that male thyroid cartilage is approximately 20% larger than female thyroid cartilage and Hirano's (1983) findings that the membranous length of the vocal folds grows at the rate of 0.7 mm per year for males and 0.4 mm for females. When fully developed, the male membrane is approximately 16 mm, while it is 10 mm for females. Titze (1989) observed the relationship between the growth of the membrane and fundamental frequency, which shows an inversed relationship: the larger the membrane, the lower the fundamental frequency. This fact has been summarized in Simpson's (2009) work regarding phonetic differences between male and female speech, noting that male speakers tend to exhibit a lower fundamental frequency than female speakers due to thicker and longer vocal fold membranes. Furthermore, this study also mentioned that the vocal tract length differs between males and females, which significantly affects the F0 of male and female speaker. Accordingly, the adult male vocal tract has an average length of 17-18 cm, whereas the adult female vocal tract has an average length of 14.5 cm. Thus, male vocals have a lower pitch than female voices due to lower resonance frequencies caused by this difference in vocal tract length. Therefore, it is widely agreed that speaker's fundamental frequency is the most robust factor helping in the identification of speaker's sex from their own voices (Childers & Wu, 1991; Crystal, 1975; Wu & Childers, 1991)

However, irrespective of the physiology, it remains controversial whether males' use of F0 really differs from females'. Crystal's (1975) observation suggested that crucial suprasegmental differences between the actual speech of males and females in the majority of languages probably exist, despite little investigation. Some studies believe that melodic and fluctuating speech are characteristics associated with typical female speech (Crystal, 1975; Wu & Childers, 1991). Moreover, Austin (1965) also mentioned that from the view of folk linguistics (non-phonetic and experimental), females' use of F0 in utterances is believed to be "swoopy." Crystal (1975) also mentioned anecdotal observations of these characteristics, such as clucking like an old hen, glissando effects, complex tone usage, and a simpering voice. As a result, there are attempts to find the acoustic correlates of this type of feminine speech. Generally, the results from many studies concerning the differences between males' and females' utterance F0, irrespective of their intrinsic F0, vary, and a strong acoustic conclusion does not exist yet (mainly due to research design and the object language, see methodology and discussion).

Aside from its phonetic function, pitch can also serve as a clue to phonological distinctions in various languages. In Thai, the target language of this study, different realizations of F0 at the lexical level lead to distinct word meanings. Thai contrasts five lexical tones (Naksakul, 2016), as shown in Table 1. Mid, low, and high tones are level tones, whereas falling and rising tones are contour tones. Moreover, pitch in Thai can be used at the interactional level in addition to its use as lexical tones. According to Luksaneeyanawin's (1998) study on Thai intonation, two key factors influence how people modify their pitch: (1) grammatical cues and (2) attitudinal cues. This aligns with Abramson's (1979) research on the impact of lexical tone on sentence prosody in Thai. They showed that the final word in a sentence—typically a grammatical word or sentence particle—indicates the sentence prosody in non-emotive speech, or speech devoid of attitudinal cues.

Table 1
Lexical Tones in Thai

Tone	Word	Transcription	Gloss
Mid	คา	/k^hāː/	'To stuck'
Low	ข่า	/k^hàː/	'galangal'
Falling	ฆ่า	/k^hâː/	'To kill'
High	ค้า	/k^háː/	'To trade'
Rising	ขา	/k^hǎː/	'leg'

Expanding on Abramson's research, Luksaneeyanawin (1998) discovered that, when compared to statements, the pitch range and level in yes-or-no questions and incomplete utterances are, respectively, narrower, and higher. Furthermore, they found that emotional cues or attitudinal meanings—such as emphatic, angry, pleasant, or surprised—also have a significant influence on how Thai sentences sound. These findings are based on Uldall's (1972) theory on intonational meanings. However, despite a slight alteration in the realization of tones in sentences, the intonation system has no effect on the phonological tones. These two studies show that pitch in Thai, particularly pitch at the utterance level, can indicate the speaker's attitude meaning. Pitch can also be seen as a linguistic variable that reflects the speaker's construction of identity.

Therefore, adapting from Luksaneeyanawin's (1998) findings, situational context is also a significant factor affecting one's modification of the fundamental frequency in utterance. Many studies indicate a relationship between pitch and the type of question being asked (Gósy & Terken, 1994; Sicoli et al., 2015). Additionally, some findings reveal the interaction of imperative utterances with realized F0 as well (Sun et al., 2008).

Thus, this research intends to study and compare the use of fundamental frequency by male and female speakers in different situational contexts. Both gender and situation, according to acoustic and perceptual views, tend to have a relationship and interaction with the speaker's fundamental frequency.

Literature Review

As aforementioned, several studies have been conducted to investigate this issue. Early studies mainly focused on intonational differences, such as Eble's (1972), which detected certain intonational patterns predominating in female speech, and Pellowe and Jones' (1978), which suggested that female speakers exhibited more rising tones than their male counterparts.

The first attempt to find the acoustic correlate of these stereotypes was made by Aronovitch (1976). A perception and evaluation test was conducted by recording connected speech obtained through a picture description task, followed by a rating task with polar adjectives (e.g., dominant/submissive). Intensity and fundamental frequency, along with their variance, were used by his participants to categorize speakers as either male or female. It is important to note that the main purpose of this research was to find the relationship between personality judgments and the speaker's sex as inferred from speech. Therefore, the correlates obtained were merely "by-products" of Aronovitch's work. Nevertheless, this study paved the way for other researchers to further investigate these correlates and the differences between male and female speech in terms of pitch realized at the utterance level.

Woods (1992) subsequently studied sex-specific pitch and intonational features in British English. Speech data were collected from 10 male and 10 female informants (five 6-7 years old speakers, and another five 27-32 years old speakers). The findings significantly show that female speakers prefer to use fall-rise and high-fall utterance F0, whereas male speakers prefer level utterance F0. This pattern is consistent in two speech settings: conversation and interview. Snidecor (1951), on the other hand, discovered that males exhibit a greater functional pitch range than females. He found that female speakers possess a slower rate of pitch change than male speakers. Tielen (1989) investigated this issue with a different object language compared to previous studies: Dutch. Ten male and 10 female speakers aged between 40-45 years were asked to produce 13 short sentences and spontaneous speech. The results showed that although male speakers possess a wider range of F0 (13 semitones), in contrast to female speakers (10 semitones), female speakers' utterance F0 has significantly larger variance (calculated by standard deviation [SD]) than male speakers. Henton (1995) found, surprisingly, no difference in male and female use of pitch in utterance, mainly due to the research design: the recordings used for analysis were from another research project about nasality in English and French.

According to previous literature, there are three primary conclusions regarding whether the fundamental frequency (F0) of male speech differs from that of female speech: (1) female speakers exhibit more fluctuation in F0 during utterances; (2) male speakers exhibit more fluctuation in F0 during utterances; and (3) there is no significant difference in F0 fluctuation between male and female speakers during utterances. It is important to consider the Just Noticeable Difference (JND) in F0 perception when interpreting these findings. JND refers to the smallest change in a stimulus that can be detected by a listener. In the context of F0, this means that even if there are measurable differences in pitch fluctuation between males and females, these differences may not be perceptible to listeners if they fall below the JND.

Moreover, these studies selected different variables—different acoustic correlates—as tools to find the answer, which is another main reason why their results are highly controversial. However, there is a tendency in later works (Henton, 1995; Tielen, 1989) to hypothesize the variance and dynamism of pitch as primary correlates of the folk linguistic description of female speech. By the term pitch variance, which Tielen (1989) used to show the difference between utterance F0 in male and female speech, it is well understood that this refers to the SD of different pitch values in one utterance. Pitch dynamism, according to Henton (1995), is defined as the change in F0 over a specific amount of time (30 milliseconds). For instance, if the F0 of utterance A measured at 1 millisecond is 235 Hz and the F0 measured at 30 milliseconds is 200 Hz, the pitch dynamism of utterance A would be 35 Hz. It should be noted that Henton converted hertz to semitones in her work. While this method of measuring pitch dynamism considers time as a factor, it is, according to Henton (1995) herself, a coarse way of obtaining pitch dynamism, posing limitations on her study. This is because there is a possibility of drastic, abrupt, or sudden pitch rises or falls occurring between 1 and 30 milliseconds. The F0 difference between 1 and 30 milliseconds does not fully represent the ongoing dynamism of the F0 throughout the entire utterance.

Regarding situational contexts that affect a speaker's modification of F0, Brinckmann and Benzmüller (1999) investigated the relationship between utterance type and F0 contour in German. Their findings suggested that the F0 in both wh-questions and declarative questions fluctuates more, with a higher range between F0 onset and F0 offset, compared to statements or yes/no questions. In contrast, Sun et al. (2008) studied pitch in imperative sentences in Standard Chinese and found that, aside from being raised compared to declarative sentences, the pitch range of imperative sentences in Standard Chinese tends to be compressed, implying a different usage of F0 in another specific context. From the two studies above, the idea that situational context is another important factor determining the projection and realization of F0 in a specific utterance cannot be dismissed.

Research Questions

Due to the gap in previous research, which failed to clearly identify the acoustic correlates of folk linguistic and stereotypical descriptions of female speech and could not quantitatively demonstrate whether a difference in F0 usage between male and female speech exists, this study aims to (1) examine whether male speech differs from female speech in terms of pitch, employing acoustic values and measurement methods designed for this purpose, and (2) explore whether the relationship between context and F0 realization in utterances exists. The study focuses on Thai, a tonal language, with the goal of attempting to offer an interesting comparison to previous research primarily conducted on stress languages such as English, French, and Dutch, to determine whether tonal characteristics influence F0 in utterances.

Methodology

Participants

The voice data were recorded from six humanities students, three males and three females, aged between 20-22 years old. All participants were native Thai speakers without any speaking problems or disorders.

Stimuli

Each of them was asked to read five stimuli sentences with no repetition, as if they were literally speaking those sentences. Each stimuli sentence, which typical Thai university students are likely to use in their daily life, represented a conversational context: affirmative, question, negation, imperative, and request. Note that, in order to create a significant distinction, wh-questions were selected rather than yes/no questions (see literature review). After finishing the stimuli sentences, they were asked to describe a picture to obtain simultaneous speech data. The total number of tokens will be six participants x six stimuli (five stimuli sentences + one connected speech) = 36 tokens.

Regarding the control of the number of words, syllables, and tonal distributions between the sentences, as well as repetitions, it should be noted that no such control was implemented in this study. This decision was made to maintain the naturalness of the speech samples. The primary goal was to capture authentic speech patterns without imposing artificial constraints that might influence the natural flow of the utterances. While this approach may introduce variability, it ensures that the data more accurately reflect spontaneous speech. Thus, the average F0 for each speaker eliciting each tone was not measured.

Moreover, this study did not emphasize lexical tone but rather focused on the long-term F0 in utterances, which is automatically normalized through the measurement of coefficient values. Lexical tones in connected speech are coarticulated and have a minimal effect on long-term F0. Nevertheless, an effort was made to control the number of syllables in the stimuli sentences. The affirmative, question, and negation sentences have a relatively similar number of syllables (8-9), whereas the imperative and request sentences have a similar number of syllables (13-14). The words in the stimuli were also controlled to be monosyllabic. Both the stimuli sentences and connected speech were analyzed for pitch dynamism (see measurement and variables sections).

Table 2

Stimuli Sentences, Alongside their Phonetic Transcription and Translation

Sentence type	Stimuli (In Thai script and phonetic transcription)	Translation
Affirmative	วันนี้กูมีเรียนตอนบ่ายโมง wān ní: kū: mī: riān tɔ̄:n bà:j mō:ŋ	‘Today I have a class at 13.00’
Question	พวกมึงจะไปนั่งที่ไหนกัน p^{h}uâk mɯ̄ŋ tɕàʔ pāj nâŋ t^{h}î: nǎj kān	‘Where are you guys going to sit (and do something)’
Negation	กูคงไม่ได้ไปกินข้าวด้วยนะ	‘I could not go to eat with you’

	kū: kʰōŋ mâj dâj pāj kīn kʰâ:w duâj náʔ	
Imperative	มึงต้องส่งงานมาให้กู ไม่งั้นกูทำต่อไม่ได้ mɯ̄ŋ tɔ̂ŋ sòŋ ŋā:n mā: hâj kū: mâj ŋán kū: tʰām tɔ̀: mâj dâj	‘You must send me the works. Otherwise, I can’t continue doing it’
Request	มึง หารกันได้ปะ กูว่ากูกินจานนี้ไม่หมด mɯ̄ŋ hǎ:n kān dâj pàʔ kū: wâ: kū: kīn tɕā:n ní: mâj mòt	‘Can we share this thing? I think I couldn’t finish it by myself’

Figure 1
Illustration of Elicitation of Simultaneous Speech (Jongsukkijpanich, 2021)


Recordings

The recording was made in a silent room using a Zoom H1N microphone covered with a windscreen, placed on a microphone stand. The distance from the microphone to the speaker’s mouth was approximately 15 centimeters. Note that due to the ongoing COVID-19 situation in April 2022 in Bangkok, Thailand, participants were required to wear face masks during the recordings to protect themselves from the virus, as people usually refused to take off their face masks unless it absolutely necessary. Research has shown that wearing face masks can have a slight effect on acoustic speech parameters, such as formant frequencies and center of gravity (Corey et al., 2020; Toscano & Toscano, 2021). However, these effects are generally minimal and do not significantly impact the reliability of F0 measurements. The sampling rate was set to 41,000 Hz, and the recordings were directly captured in Audacity. Participants were asked to speak each sentence as naturally as possible, pause for about 5-7 seconds, and then start speaking another sentence. After recording the read speech, the task description for recording simultaneous speech was given orally by the researcher, and then the stimulus picture was shown to the participant.

Measurement

The recordings were subsequently imported into PRAAT (Boersma & Weenink, 2022). They were later segmented to the points where pitch begins and stops being evident. Within the simultaneously connected speech, the initial utterance, which aligns with the definition of an utterance—a speech with a substantial pause at the beginning and end (approximately 5 seconds)—was selected. The first and last 10% of the segmented utterances were then excised. Following this, the fundamental frequency (F0) of each utterance was measured every 1 millisecond using the pitch listing feature in PRAAT. Any undefined F0 values were replaced with 0. A MATLAB script was utilized to plot the values on a graph, generate a trendline, and compute the coefficients (perturbation), SDs (variance), and other relevant metrics.

Variable

From the literature review, I selected pitch dynamism (Henton, 1995) as an acoustic measure to analyze F0 usage in utterances. As defined, pitch dynamism captures the variation in pitch over time, revealing how speakers modulate their F0 throughout an utterance. However, as noted previously, it lacks the precision needed to fully capture these pitch changes. Fuchs (2018) even proposed using pitch variance as an alternative definition of pitch dynamism to address this limitation. While the SD of utterance F0 can also reveal how speakers modulate pitch, it loses the temporal dimension, which is crucial for understanding F0 usage within an utterance.

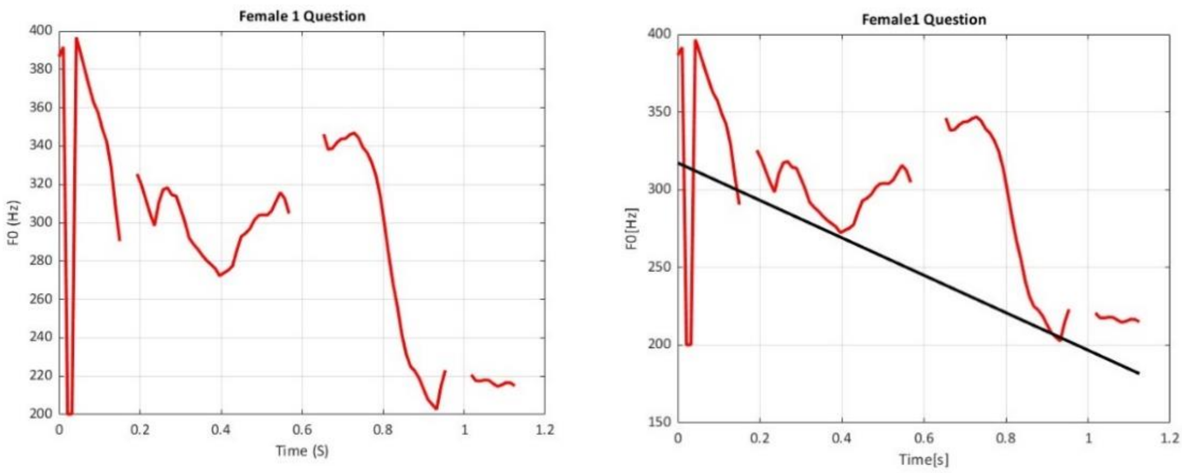
Since pitch dynamism cannot be solely determined by the change in pitch over a specific time limit, I propose combining both aspects of dynamism, time, and variance, as used in previous literature. This redefines pitch dynamism with the following definition: pitch dynamism is a two-dimensional coordinate value (x, y) comprised of (1) pitch perturbation (X-axis) and (2) pitch variance (Y-axis).

Pitch perturbation can be defined as the rate and directionality of F0 change over time. It is obtained by measuring the coefficient value/slope from the linear equation $y = ax + b$ created by plotting a trendline through an utterance F0 plot, which is generated by measuring F0 every 1 millisecond. The rate of F0 change is represented by the absolute coefficient value: a higher value indicates a faster rate. The directionality is then represented by the polarity of the coefficient value: positive for rising pitch, negative for falling pitch.

Pitch variance can be defined as the spread of the F0 value within a specific utterance. It is obtained by calculating the SD from the F0 measurement: a higher value indicates greater fluctuation. While pitch perturbation primarily addresses the F0 movement in an utterance, including its directionality, pitch variance focuses on the spread of F0 values. With these two values, pitch dynamism can be plotted on a graph. This method and definition of pitch dynamism, compared to Henton's (1995), are capable of capturing not only the fine-grained pitch movement and directionality over time, but also the variance of the pitch within a specific utterance.

For the statistical and quantitative analysis, t-test and f-test were used to determine whether the different between pitch perturbation and pitch variance, respectively, in male and female speech was statistically significant. Additionally, ANOVA was used to investigate the interplay between the gender variable and situational variable.

Figure 2
Sample F0/time Plot without (right) and with (left) Linear Trendline of a Question Sentence



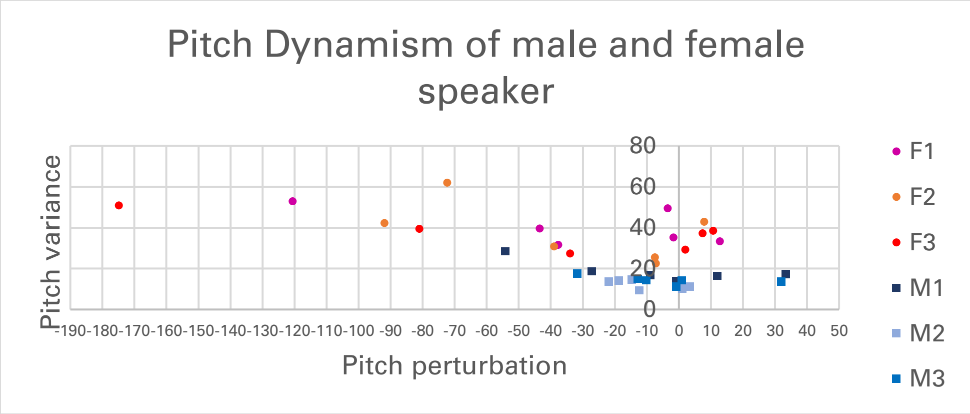
Results

Gender Factor

The plot in Figure 3 provides a visual representation of pitch dynamism for both male and female speakers, specifically examining their pitch variance and perturbation, addressing the first research question.

Most male speakers exhibit a pitch variance below 20, which is relatively low compared to female counterparts, indicating that the spread of their F0 values within a specific utterance is relatively narrow. This suggests that male speakers tend to maintain a more stable pitch range, with less fluctuation in their voice. Additionally, the pitch perturbation for male speakers primarily ranges between -40 and 20. This range demonstrates that the rate and directionality of their pitch changes are generally less extreme. An exception to this pattern was observed with Male Speaker 1 during a question, where the pitch perturbation was higher. This anomaly can be attributed to the situational context.

Figure 3
Plotted Pitch Dynamism of Male and Female Speaker from all Tokens (Both Elicited Read and Spontaneous Connected Speech)



In contrast, female speakers show a wider range of pitch variance, with all of the values extending beyond 20, suggesting variability of the pitch within utterances and a greater range of vocal modulation. Furthermore, female speakers display a broader spectrum of pitch perturbation, ranging from -180 to 50. This broader range suggests a more dynamic use of pitch, encompassing both rapid and varied directionality of F0 changes.

The data clearly demonstrate distinct differences in how male and female speakers use F0 in their utterances. Male speakers tend to have a more constrained pitch variance and perturbation, indicating a more stable and less fluctuating pitch. This stable pitch usage in males contrasts with the greater pitch dynamism observed in female speakers, who exhibit both higher pitch variance and broader pitch perturbation.

By analyzing the pitch variance and perturbation, it becomes evident that female speakers engage in a more dynamic vocal pattern compared to their male counterparts. This dynamic pattern is characterized by rapid and varied changes in pitch directionality and rate, which contribute to a more fluctuating and melodious speech quality. On the other hand, male speakers' more constrained pitch modulation reflects a steadier and less variable vocal pattern.

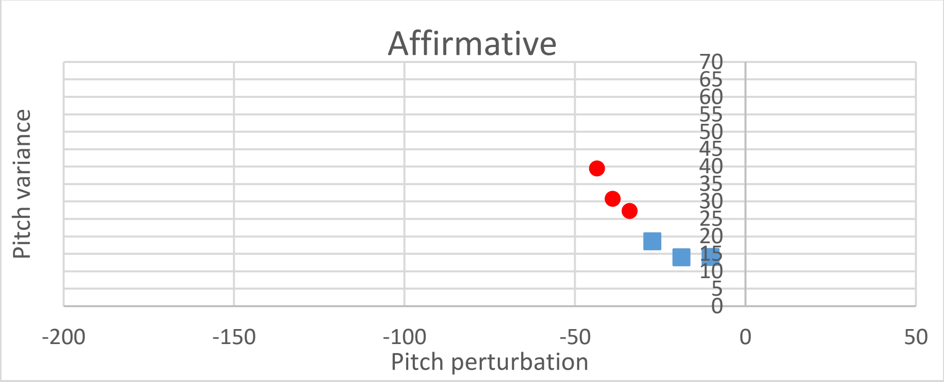
These observations provide a foundational basis for understanding the acoustic correlates of gender differences in speech. The broader pitch variance and perturbation in female speech are consistent with the stereotype that female speakers often have a more melodious sound.

Situational Factor

The situational factor revealed distinct patterns of pitch variance and perturbation based on the type of utterance. Figures 4-9 illustrate these patterns for affirmative, question, and negation contexts, with red dots representing female speakers and blue dots representing male speakers. In affirmative utterances (Figure 4), both male and female speakers show relatively low pitch variance and perturbation. Female speakers, however, display slightly higher values in both dimensions compared to their male counterparts. This indicates a tendency for female speakers to use a somewhat more dynamic pitch range even in affirmative contexts, but the overall dynamism is still restrained.

Figure 4

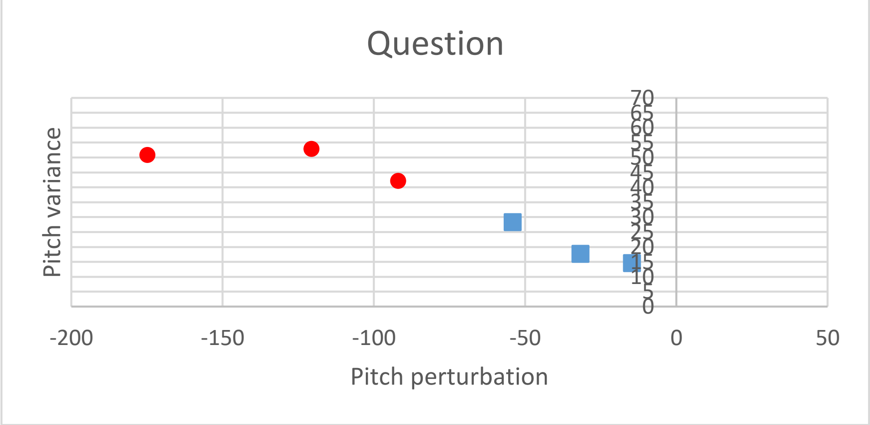
Plotted Pitch Dynamism of Male and Female Speakers in Affirmative Context



Question utterances (Figure 5) exhibit the most distinctive pattern. Female speakers show a broad range of pitch perturbation, extending significantly into the negative direction. This suggests that questions prompt a considerable lowering of pitch among female speakers, potentially as a strategy to convey uncertainty or inquisitiveness. Male speakers also show a noticeable pitch perturbation, but the range is less extreme compared to females. This highlights a situational context where the dynamic use of pitch is markedly influenced by gender, with females demonstrating greater variability.

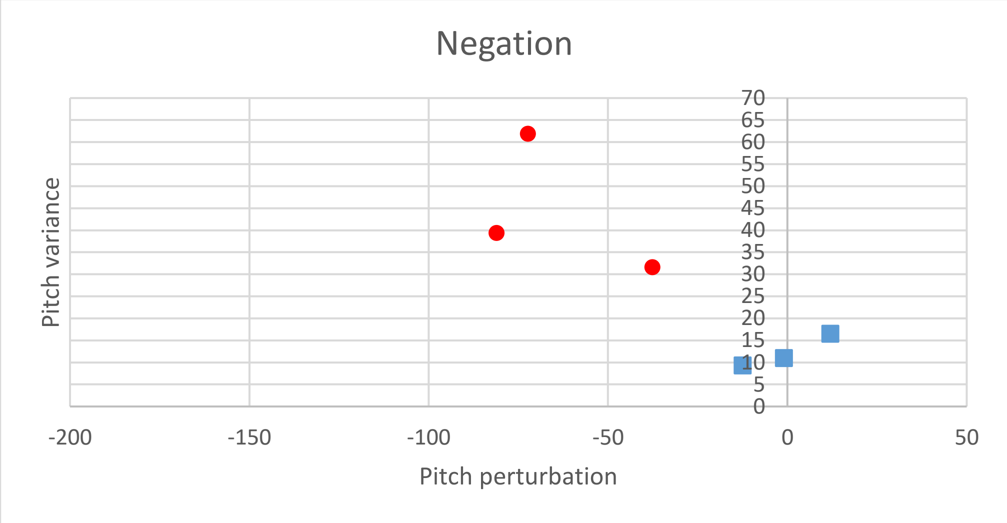
Figure 5

Plotted Pitch Dynamism of Male and Female Speakers in Question Context



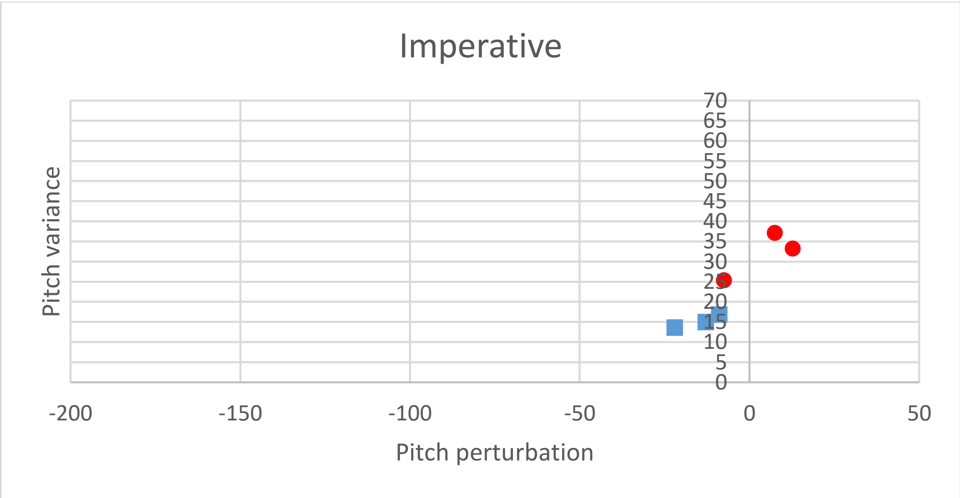
Negation utterances (Figure 6) present a diverse combination of pitch perturbation and variance. Female speakers, in particular, exhibit wide-ranging pitch perturbation and pitch variance, indicating a more pronounced modulation in their speech.

Figure 6
Plotted Pitch Dynamism of Male and Female Speakers in Negation Context



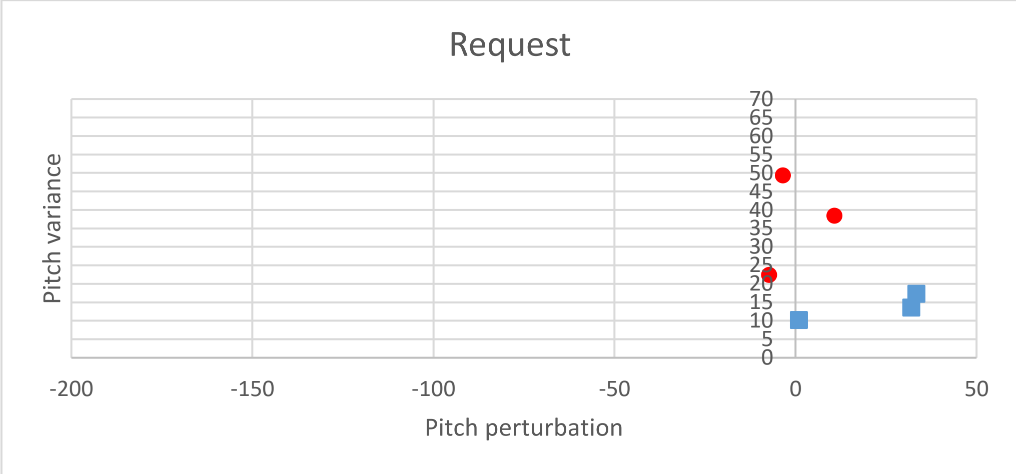
In imperative utterances (Figure 7), there is an interspeaker variation in pitch variance, predominantly influenced by gender. Although female speakers show higher pitch variance, the pitch perturbation remains close to zero for both genders, indicating that while the pitch range varies, the overall direction of pitch change is stable.

Figure 7
Plotted Pitch Dynamism of Male and Female Speakers in Imperative Context



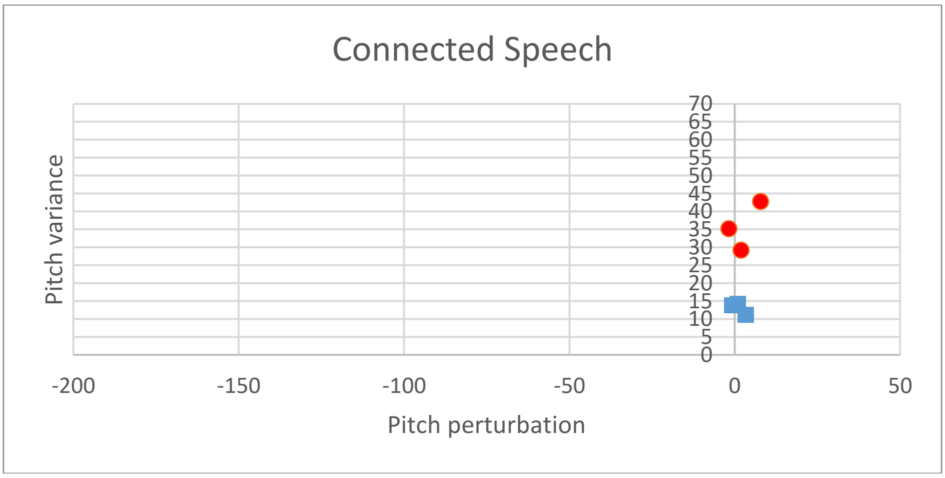
Request utterances (Figure 8) show a pattern similar to imperatives, with female speakers exhibiting higher pitch variance. Surprisingly, male speakers tend to have a much more fluctuating pitch perturbation in comparison to female counterparts. This could be due to the polite and softer tone typically used in requests, where abrupt pitch changes might be less frequent.

Figure 8
Plotted Pitch Dynamism of Male and Female Speakers in Request Context



In the connected speech, female speakers generally have higher pitch variance and a slightly wider range of pitch perturbation, indicating more dynamic and varied use of pitch in connected speech, whereas male speakers tend to have lower pitch variance and more restrained pitch perturbation, reflecting a more stable and consistent pitch usage. These patterns align with the initial observations from the study, suggesting that gender significantly influences pitch usage, with female speakers exhibiting greater vocal modulation compared to male speakers. Note that the small range of pitch perturbation here results from connected speech mainly containing affirmative sentences.

Figure 9
Plotted Pitch Dynamism of Male and Female Speakers in Connected Speech



Analysis

Figure 9 displays pitch dynamism by gender, and the fact that female speakers exhibit a wider range of pitch dynamism compared to males is demonstrated visually. Subsequently, statistical tests were conducted to confirm the findings. A Welch's T-test confirmed a significant difference in pitch perturbation between genders (p=0.033792). Further analysis with a two-sample F-test for variances (p=0.000261) revealed a significantly greater difference in pitch variances between genders. The much lower p-value of the F-test suggests that pitch variance is a stronger cue than pitch perturbation for distinguishing male from female speech.

This study's findings support Tielen's (1989) claim that SD is a key indicator to distinguish between male and female utterance F0, independent of physiological factors. The higher pitch variance observed in female speakers aligns with the stereotype of a more dynamic and melodic speech pattern. Conversely, male speakers exhibit a more stable and less variable pitch.

Table 3

T-Test for Pitch Perturbation Difference

T-Test: Two-Sample Assuming Unequal Variances

	Female	*Male*
Mean	-37.4248	-7.2994
Variance	2742.964	444.6502
Observations	18	18
Hypothesized Mean Difference	0	
df	22	
t Stat	-2.26379	
P(T<=t) one-tail	0.016896	
t Critical one-tail	1.717144	
P(T<=t) two-tail	**0.033792**	
t Critical two-tail	2.073873	

Table 4

F-Test for Pitch Variance Difference

F-Test Two-Sample for Variances

	Female	*Male*
Mean	38.26895	14.95684
Variance	108.7332	17.77263
Observations	18	18
df	17	17
F	6.118015	
P(F<=f) one-tail	0.000261	
F Critical one-tail	2.271893	

Unlike the clear patterns observed in other analyses, the situational variation plots were inconclusive, making it difficult to directly compare F0 usage across different contexts. To address this, separate ANOVAs for pitch perturbation and pitch variance were conducted. The ANOVA results revealed that gender alone has no significant effect on F0 usage. However, situational context and the interaction between gender and context were found to have a significant impact. This suggests that F0 usage differs between males and females primarily in specific situational contexts, highlighting the importance of considering context when analyzing pitch dynamism.

Table 5

ANOVA Analysis of Pitch Perturbation

Analysis of variance (ANOVA)

Source	*Sum Sq.*	*d.f.*	*Singular?*	*Mean Sq.*	*F*	*Prob>F*
Gender	722.6785	1	0	722.6785	2.05548	0.167109
Situation	29417.77	4	0	7354.443	20.91789	**0.000000648**
Gender*Situation	20861.99	4	0	5215.499	14.8342	**0.00000879**
Error	7031.724	20	0	351.5862		
Total	58034.17	29	0			

Further analysis using pitch variance ANOVA shows that gender is a significant factor influencing pitch, while situational context alone is not. This suggests that pitch variance is the key indicator of speaker gender, whereas pitch perturbation primarily reflects the speaker's situation.

Table 6

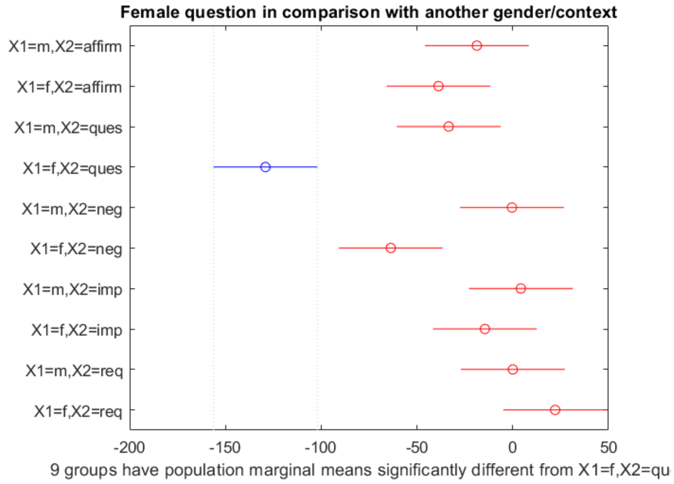
ANOVA Analysis of Pitch Variance

Analysis of variance (ANOVA)

Source	*Sum Sq.*	*d.f.*	*Singular?*	*Mean Sq.*	*F*	*Prob>F*
Situation	2881.28	1	0	2881.2754	48.50	0.170123
Gender	518.5953	5	0	103.7191	1.75	**0.000000926**
Gender*Situation	278.8163	5	0	55.7633	0.94	0.477483
Error	1188.171	20	0	59.4085		
Total	5462.342	36				

The visualization of ANOVA results highlights the distinctiveness of female questions, indicating that this context particularly stands out, a topic that will be explored further in the discussion section.

Figure 10
ANOVA Visualization of Female Question in Comparison with Other Contexts



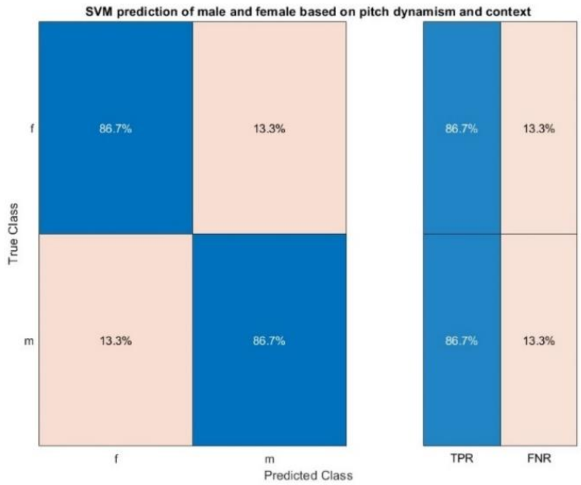
Overall, the analysis confirms that pitch dynamism differs significantly between male and female speakers, with pitch variance being a more critical factor than pitch perturbation, answering the first research question. Situational context plays a crucial role in these differences, suggesting that F0 usage in speech is context dependent. The significant gender-context interaction further underscores the complexity of pitch dynamism and the necessity of considering situational factors in gender-based speech analysis, answering the second research question.

Discussion

Machine Learning Approach

According to several previous studies (Childers & Wu, 1991; Henton, 1995; Wu & Childers, 1991), computational linguists have also explored the topic of distinguishing male speech from female speech using fundamental frequency as a primary distinguishing factor. Given that the method presented in this work yields significant and distinct results in separating male from female speech in specific contexts using redefined pitch dynamism, I employed a Support Vector Machine (SVM) model based on the pitch perturbation ANOVA. SVM is a supervised machine learning model that is particularly effective for classification tasks, as it seeks to find the optimal hyperplane that maximizes the margin between different classes. The SVM model used here demonstrated an 86.7% accuracy in predicting gender based on pitch dynamism, with a 13.3% error rate. This level of correctness supports the findings, with the error rate potentially attributable to the exception mentioned earlier (Male speaker 1, Question).

Figure 11
Result of SVM Prediction Based on Pitch Perturbation ANOVA



Interaction of Context: Explanation of Pitch Perturbation in Different Contexts

The results show a general spread of pitch dynamism in both affirmative and negative utterances, as well as in utterances segmented from connected speech. This is likely due to the generic nature of these utterances. Interestingly, based on the results, negation appears to be a sub-category of affirmation, as both serve to convey facts or opinions. All utterances segmented from connected speech were classified as affirmative sentences.

One male speaker (M1 Question) presented as an outlier, with pitch dynamism falling within the typical range for females. This speaker's F0 exhibited greater perturbation and fluctuation compared to the other male speakers. The situational context likely explains this outlier. Previous research by Brinckmann and Benzmüller (1999) suggests that wh-questions tend to have more F0 fluctuation than other question types and affirmative sentences. Since the stimulus was a wh-question, it is unsurprising that this utterance showed the most pitch fluctuation among the participant's recordings. This is further supported by the finding that the wh-question stimulus elicited the most F0 perturbation in all female participants as well.

The ANOVA visualization shows that the female question is the only one clearly distinct from all others. This suggests that female speakers naturally exhibit more fluctuation and melodic variation in their fundamental frequency (F0) during utterances, and question sentences allow for the greatest variation in F0. These factors likely explain this phenomenon.

Another interesting context to discuss is the imperative. From the general results, a low dispersion of F0 and alignment to zero in terms of perturbation can be observed. Although there are no direct studies on Thai intonation patterns of imperative sentences, the results of this study partly correspond to the findings from Sun et al. (2008), which suggested that the pitch range of imperatives in Chinese, a tonal language, is likely to be compressed. Similar characteristics might be inferred for Thai due to the tonal nature of both languages, but further research specific to Thai is necessary for a more accurate comparison.

The request context generally exhibits a wide dispersion of pitch dynamism. Here, it cannot be compared directly with the general dispersion in affirmatives, as I consider the request context to be a combination of both question and imperative contexts. The high dispersion, combined with low dispersion, results in an overall general dispersion.

Interaction of Tone and Pitch Dynamism

The results of this research differ from previous studies, such as those by Tielen (1989) and Henton (1995), in several aspects. For instance, Tielen's (1989) study yielded significant results with the SD but with rather low significance. Henton's (1995) study, on the other hand, yielded insignificant results regarding pitch dynamism. I suggest that the differences between the results of these previous studies and the current work can be attributed to the interaction between phonological tone and pitch dynamism (F0) throughout the utterance.

According to Eady (1982), F0 patterns in tone languages, as opposed to stress languages, are influenced by the tone contours of each lexical item in the utterance, resulting in more fluctuating F0 at the utterance level. The previous studies reviewed focused on stress languages such as Dutch (Tielen, 1989) and French and English (Henton, 1995). In contrast, this study investigates Thai, a tone language. Therefore, the significant results regarding pitch variance in this study, compared to the less significant results in previous works, suggest that pitch variance is a primary cue in tone languages. Furthermore, the less significant results for pitch perturbation in this study imply that F0 modification in utterances is language-specific rather than cross-linguistic. Each language has its unique way of modifying F0 throughout the utterance based on the gender of the speaker.

Limitations

This study has several limitations. Firstly, the sample size is small (n=6; m=3, f=3), and the presented demographic factors such as age are also constrained, which limits the generalizability of the findings. It serves as a preliminary case study aimed at confirming the hypothesis that there are differences in the fundamental frequency (F0) between male and female utterances, thereby providing a foundational basis for the proposed research method of measuring utterance F0 using an adapted version of pitch dynamism.

Additionally, the number of tokens (n=36) is relatively low, as efforts were made to ensure each elicitation sounded as natural as possible. Despite these limitations, the results obtained still reflect ongoing patterns in the characteristics of F0 in utterances by both male and female speakers across different situational contexts.

Conclusion

In conclusion, the findings show that variations in the modification of F0 throughout an utterance in Thai exist between male and female speech, with situational context being another main factor influencing these variations, as suggested by the statistical analysis. Furthermore, there is a connection between the type of utterance, its communicative function, and the realized F0 from the speaker. Pitch variance serves as the main cue to distinguish gender, while pitch perturbation serves as the main cue to distinguish conversational context. This study is significant as it addresses the question of male and female F0 differences beyond physiological factors and adds new insight to the ongoing discussion. Additionally, this study paves the way for further research on pitch variation in various utterances, both in the domain of the Thai language and in general linguistic studies. I suggest that future studies investigate the relationship between pitch dynamism and gender identity more deeply, especially in non-normative contexts, as this could enhance our understanding of the construction of gender identity through acoustic correlates.

Author

Athit Wu was born in 1999, in Bangkok, Thailand. He completed his undergraduate studies at the Faculty of Arts, Chulalongkorn University, majoring in German with a minor in Linguistics under the Honors Program. Following his bachelor's degree, he continued his academic journey by enrolling in the Master of Arts in Linguistics program at the same institution. Athit's research interests lie in the field of Sociolinguistics, Sociophonetics, and Phonology.

Acknowledgements

Foremost, I would like to express my sincere gratitude to Asst. Prof. Dr. Sujinat Jitwiriyanont for the continuous support, for his patience, motivation, enthusiasm, immense knowledge, and advice. Additionally, I would like to thank Asst. Prof. Dr. Pavadee Saisuwan for her support, for her advice concerning my projects, and for her positive energy. My sincere thanks also go to Francesco Burroni for the suggestions and comments on statistical and machine learning approaches, as well as Gena Jongsukkijpanich for her illustration. My further thanks also go to all of the informants who participated in this study. Finally, I would like to thank anonymous reviewers and editor teams of NIDA Journal of Language and Communication for their valuable suggestions and comments.

References

Abramson, A. S., Fischer-Jørgensen, E., Rischel, J., & Thorsen, N. (1979). Lexical tone and sentence prosody in Thai. *Haskins laboratories: Status Report on Speech Research SR-59, 60*, 153-160.

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology, 99*(2), 207-220. https://doi.org/10.1080/00224545.1976.9924774

Austin, W. M. (1965). Some social aspects of paralanguage. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *11*(1), 31-39. https://doi.org/10.1017/S0008413100005703

Brinckmann, C., & Benzmüller, R. (Eds.). (1999). *Proceedings, Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99), Budapest, Hungary, September 5-9, 1999*. https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/6860

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer [Computer program]. Version 6.2.13. http://www.praat.org/

Childers, D. G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *The Journal of the Acoustical society of America*, *90*(4), 1841-1856. https://doi.org/10.1121/1.401664

Corey, R. M., Jones, U., & Singer, A. C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America*, *148*(4), 2371-2375. https://doi.org/10.1121/10.0002279

Crystal, D. (1975). *The English tone of voice: Essays in intonation, prosody and paralanguage*. Hodder & Stoughton Educational.

Eady, S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech*, *25*(1), 29-42. https://doi.org/10.1177/002383098202500103

Eble, C. C. (1972). How the speech of some is more equal than others. Paper presented at Southeastern Conference on Linguistics. In Annotated Bibliography, B. Thorne & N. Henley (Eds.), *Language and sex: Difference and dominance*, p. 246. Newbury House, Rowley.

Eckert, P., & McConnell-Ginet, S. (2003) *Language and gender*. Cambridge University Press.

Fuchs, R. (2018). Pitch range, dynamism and level in postcolonial varieties of English: A comparison of Educated Indian English and British English. *Proceedings of the 9th Speech Prosody*, 893-897. https://doi.org/10.21437/SpeechProsody.2018-180

Gósy, M., & Terken, J. (1994). Question marking in Hungarian: Timing and height of pitch peaks. *Journal of Phonetics*, *22*(3), 269-281. https://doi.org/10.1016/S0095-4470(19)30204-9

Henton, C. (1995). Pitch dynamism in female and male speech. *Language & Communication*, *15*(1), 43-61. https://doi.org/10.1016/0271-5309(94)00011-Z

Hirano, M. (1983). The structure of vocal folds. In I. R. Titze (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*, 1699-1707.

Jongsukkijpanich, P. (2021, N.D.). *Portfolio*. Gena Jongsukkijpanich https://genajong.portfolio.site/

Kahane, J. (1978). A morphological study of the human prepubertal and pubertal larynx. In I. R. Titze (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America, 85*, 1699-1707.

Luksaneeyanawin, S. (1998). *Intonation in Thai.* (D. Hirst and A. D. Christo, Eds.). Cambridge University Press.

Naksakul, K. (2016). The sound system of Thai language [ระบบเสียงภาษาไทย]. Bangkok: Chulalongkorn University Faculty of Arts Academic Dissemination Project

Ohala, J. J. (1984). An Ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1-16. https://doi.org/10.1159/000261706

Pellowe, J., & Jones, V. (1978). *On intonational variability in Tyneside speech.* (P. Trudgil, Ed.). Arnold.

Sicoli, M. A., Stivers, T., Enfield, N. J., & Levinson, S. C. (2015). Marked initial pitch in questions signals marked communicative function. *Language and Speech*, *58*(2), 204-223. https://doi.org/10.1177/0023830914529247

Simpson, P. A. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621-640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Snidecor, J. C. (1951). The pitch and duration characteristics of superior female speakers during oral reading. *Journal of Speech and Hearing disorders, 16*(1), 44-52. https://doi.org/10.1044/jshd.1601.44

Sun, J., Lu, J., Li, A., & Jia, Y. (2008). The pitch analysis of imperative sentences in Standard Chinese. In *2008 6th International Symposium on Chinese Spoken Language Processing* (pp. 1-4). IEEE.

Tielen, M. T. (1989). Fundamental frequency characteristics of middle aged men and women. *Proceeds of the Institute of Phonetic Sciences Amsterdam*, *13*, 49-58.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*, 1699-1707. https://doi.org/10.1121/1.397959

Toscano, J. C., & Toscano, C. M. (2021). Effects of face masks on speech recognition in multi-talker environments. PLoS ONE, 16(2): e0246842. https://doi.org/10.1371/journal.pone.0246842

Uldall, E. (1972). Dimensions of meaning in intonation. In *Readings in modern linguistics: An anthology* (pp. 232-239). De Gruyter Mouton. https://doi.org/10.1515/9783110820041-015

Woods, N. (Eds.) (1992*). Proceedings of the IRCS Workshop on Prosody in Natural Speech, August 5-12, 1992*. University of Pennsylvania.

Wu, K., & Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis. *The Journal of the Acoustical Society of America*, *90*(4), 1828-1840. https://doi.org/10.1121/1.401663