DETERMINATION OF APPROXIMATE HEIGHT AND BODY MASS OF AN INDIVIDUAL BY SPECTRAL ANALYSIS OF VOICE.

Submitted By

PRAGATHI.S

22MSRFS050

Project submitted to

Department of Forensic Science

JAIN (Deemed-to-be University), Bengaluru, Karnataka

In partial fulfillment of the University Regulations for the Degree of

MASTER OF SCIENCE

IN

FORENSIC SCIENCE

UNDER THE GUIDANCE OF

MR. N VISHNU VENKATESH

Assistant Professor

Batch 2022-2024

DEPARTMENT OF FORENSIC SCIENCE

CERTIFICATE

This is to certify that the project entitled **DETERMINATION OF APPROXIMATE HEIGHT AND BODY MASS OF AN INDIVIDUAL BY SPECTRAL ANALYSIS OF VOICE** is a record of bonafide study and research carried out by **Pragathi.S**, under my supervision and guidance as the partial fulfillment of the University regulation for the course of **M.Sc in Forensic Science.**

Date :
Place : Bengaluru

Mr. Vishnu Venkatesh.

Project Supervisor
Department of Forensic Science,
JAIN (Deemed-to-be University)
Bangalore

Dr. Reena Susan P.

Head
Department of Forensic Science,
JAIN (Deemed-to-be University)
Bangalore

Dr. Asha Rajiv

Director
School of Sciences,
JAIN (Deemed to be University)
Bangalore

DECLARATION

I, **PRAGATHI.S,** hereby declare that this major project entitled **DETERMINATION OF APPROXIMATE HEIGHT AND BODY MASS OF AN INDIVIDUAL BY SPECTRAL ANALYSIS OF VOICE** is a bonafide and genuine research work carried out by me under the guidance of **MR.N VISHNU VENKATESH**. This project, in full or part, has not been submitted to this or any other University before, for the award or any degree.

Place: Bengaluru

Date:

PRAGATHI.S

22MSRFS050

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude and appreciation to **Jain University** for having deputed me for this dissertation work in the institution. I express my heartfelt gratitude to my guide **Mr. Vishnu Venkatesh** and **Mrs. Priyanksha Das,** Assistant Professor, Department of Forensic Science, Jain University, Bangalore for providing valuable insight and support leading to the successful completion of my dissertation work.

We are also grateful to **Dr. Asha Rajiv**, Director, School of Sciences, **Dr. K. N. Varalakshmi**, Director, Centre for Research in Pure & Applied Sciences and **Dr. K. R. Sridhara Murthi,** Director of Academics & Planning, JAIN (Deemed-to-be University) for providing us with this opportunity. We are thankful to **Dr. Reena Susan P.,** Head of the Department of Forensic Science for the constant support and encouragement.

We thank **Dr. Chenraj Roychand**, Chancellor, JGI, **Dr. N. Sundararajan**, Pro-Chancellor and **Dr. Raj Singh**, Vice-Chancellor, JAIN (Deemed-to-be University), for providing the required infrastructure and support without which this project would not have been possible.

I am grateful to acknowledge people who have helped me during my dissertation work by providing samples required for my research work. It’s my immense pleasure to thank my parents for their support throughout my studies.

ABSTRACT

Voice analysis, a subset of biometric identification technology, has gained significant attention for its wide-ranging applications across security, law enforcement, healthcare, and telecommunications sectors. This study delves into the science of spectral analysis of voice and its potential for estimating physical attributes such as height and body mass.

Sound waves, comprising varying frequencies, carry crucial information about the physical characteristics of the speaker. Spectral analysis allows for insightful examination of these attributes. Notably, the length of the vocal tract, strongly linked to height, and the fundamental frequency of the voice, closely associated with body mass, are of particular interest. Extending previous research, which mainly focused on adults, this study aims to estimate the height and body mass of children through spectral analysis of their voices. Fundamental frequency, a key parameter derived from spectral analysis, is utilized to predict body mass, with heavier individuals typically exhibiting lower fundamental frequencies. Through the exploration of spectral analysis techniques and their application to voice data from children, this study contributes to the broader understanding of biometric identification methods. The findings shed light on the potential of voice analysis as a non-invasive and efficient means of estimating physical attributes, with implications for various domains including pediatrics and biometric security.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Digital forensics, as a facet of forensic science, concentrates on recognizing, obtaining, processing, scrutinizing, and documenting electronically stored data. Given that electronic evidence pervades nearly all criminal undertakings, the support provided by digital forensics is indispensable for law enforcement inquiries.

1. 1 Audio forensics- Audio forensics is a specialized field within forensic science that deals with the acquisition, analysis, enhancement, and interpretation of audio recordings for investigative and legal purposes. It involves techniques and technologies to authenticate, clarify, and evaluate audio evidence such as recordings of conversations, phone calls, or other sounds that may be relevant to criminal investigations, legal proceedings, or other disputes. Audio forensics experts use a variety of tools and methodologies to examine audio recordings, including spectrographic analysis, noise reduction, voice comparison, and speaker identification, to determine the authenticity, integrity, and context of the audio evidence

1.2 Voice analysis- Voice analysis, also known as voice biometrics or speaker recognition, refers to the process of identifying individuals based on their unique vocal characteristics. Unlike other biometric modalities such as fingerprints or iris scans, voice analysis relies on the distinctive features of an individual's voice, including pitch, tone, rhythm, and pronunciation. With the advancements in technology, voice analysis has become an integral component of various applications ranging from security systems to forensic investigations.

1.2.1 Spectral analysis of voice- Spectral analysis of voice involves the examination of the frequency content of speech signals to extract meaningful information about vocal characteristics and speech production mechanisms. This paper provides an overview of spectral analysis techniques used to analyze voice signals and explores their applications in fields such as linguistics, speech pathology, and telecommunications.

1.2.2 Fundamental Concepts of Spectral Analysis: a. Fourier Transform: The foundation of spectral analysis, Fourier transform decomposes a time-domain signal into its frequency components, revealing the spectral characteristics of the signal. b. Spectrogram: A graphical representation of the time-varying frequency content of a signal, spectrograms are widely used in voice analysis to visualize vocal tract resonances, formants, and speech prosody.

1.2.3Techniques in Spectral Analysis of Voice:
a. Linear Predictive Coding (LPC): LPC analysis models the vocal tract as a linear filter to estimate formant frequencies and bandwidths, providing insights into speech articulation and vowel production.
b. Cepstral Analysis: Cepstral analysis involves transforming the logarithm of the spectrum to separate the vocal tract and excitation source components of speech, facilitating pitch detection and voice quality assessment.
c. Harmonic-to-Noise Ratio (HNR): HNR analysis quantifies the balance between harmonic components (related to voiced speech) and noise components (related to unvoiced speech or background noise), aiding in voice quality assessment and voice pathology diagnosis. (Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006))

1.3 Production of human voice- The production of the human voice involves a complex interplay of physiological structures and processes within the respiratory, laryngeal, and vocal tract systems. Here's an overview of how the human voice is produced:

1.3.1 Respiratory System-The process begins with the respiratory system, which includes the lungs and the diaphragm. When we breathe in, air is drawn into the lungs, and when we exhale, air is expelled. For speaking or singing, we regulate the airflow from the lungs to create pressure beneath the vocal folds.

1.3.2 Laryngeal System: Located in the throat, the larynx (or voice box) houses the vocal folds (also known as vocal cords). The vocal folds are comprised of mucous membranes and muscles, and they are crucial for voice production. When we speak or sing, the vocal folds come together and vibrate as air from the lungs passes through them. This vibration produces sound. (Lass N. J., 1980)

1.3.3 Phonation: The process of the vocal folds vibrating to produce sound is called phonation. The pitch of the sound is determined by the tension of the vocal folds and the rate of vibration. Higher tension and faster vibration produce higher-pitched sounds, while lower tension and slower vibration produce lower-pitched sounds. (Jaid, U. H., & Abdulhassan, A. K. (2023)

1.3.3.1 Articulatory System: The vocal tract, which includes the mouth, throat, and nasal cavity, acts as a resonating chamber that shapes the sound produced by the vocal folds. By changing the shape and position of the tongue, lips, and other articulators, we can alter the quality and timbre of the sound, producing different vowels and consonants.

1.3.3.2 Resonance: As the sound produced by the vocal folds travels through the vocal tract, it is shaped and amplified by the resonating cavities within the mouth and throat. The size and shape of these cavities affect the characteristics of the sound, contributing to the richness and clarity of the voice.

1.3.3.3 Prosody: Beyond individual sounds, the human voice also conveys meaning and emotion through variations in pitch, rhythm, and intonation. This aspect of speech, known as prosody, involves the modulation of pitch and timing to convey nuances such as emphasis, emotion, and syntactic structure. (Poorjam, (2015, March))

1.4 Praat software- PRAAT is a freeware program for the analysis and reconstruction of acoustic speech signals. You can analyze, synthesize, manipulate speech, and create high-quality pictures for your articles and thesis with it. Praat is an open software tool for the analysis of speech in phonetics. It was designed and continues to be developed, by Paul Boersma and David Weenink of the University of Amsterdam. It's free and available for most platforms. Praat was designed to cater for different needs with an easy interface, many default options to learn by trying, a searchable manual, and various possibilities of analysis, manipulation, and labeling (Goldman, 2004: 1). There are many Praat tutorials available for helping with the Praat application. However, the majority of the existing Praat manuals were designed for software documentation and assumed a strong phonetics or programming background of readers. The current manual is compiled from a variety of elaborate manuals with a special focus on those most frequently used functions and techniques for acoustic analysis. The target readers of this are those beginners who are not equipped with a strong phonetics or programming background but want to do some phonetic analysis of speech sounds. A clear visual presentation of operational procedures and an introduction to acoustic knowledge are provided to facilitate the use of Praat in linguistic research.

1.5 Pitch- Pitch refers to the perceived frequency of a sound, which is closely related to how high or low we perceive a sound to be. In simpler terms, pitch is how "high" or "low" a sound seems to be.

1.6 Working of pitch-

1.6.1 Frequency: The pitch is determined by the frequency of sound waves. Frequency is measured in Hertz (Hz), which represents the number of cycles of vibration per second. Higher frequencies result in higher-pitched sounds, while lower frequencies produce lower-pitched sounds.

1.6.2 Human Voice: In the context of the human voice, pitch is controlled by the tension and length of the vocal folds in the larynx (voice box). When the vocal folds are tensed and shorter, they vibrate more quickly, producing higher-pitched sounds. Conversely, when the vocal folds are relaxed and longer, they vibrate more slowly, resulting in lower-pitched sounds. By adjusting the tension and length of the vocal folds, we can produce a wide range of pitches, enabling us to speak and sing with different tones and melodies.

1.7 Pulse- In spectrum analysis of voice, the term "pulse" typically refers to a periodic excitation of the vocal folds. This periodic excitation is essential for generating voiced speech sounds, such as vowels and voiced consonants. When the vocal folds come together and vibrate periodically, they produce a series of pulses or glottal pulses. These pulses create a fundamental frequency, which determines the pitch of the voice. pulses in spectrum analysis of voice refer to the periodic excitation of the vocal folds, which can be observed in the frequency domain as harmonics or formants. Analyzing these pulses can provide valuable information about the voice and its production mechanisms.

1.8 Voicing- In spectrum analysis of voice, "voicing" refers to the presence or absence of periodic oscillations in the speech signal, which indicate whether the vocal folds are vibrating. Voiced sounds occur when the vocal folds vibrate, producing a periodic waveform with distinct harmonics, while unvoiced sounds occur when there is turbulent airflow through the vocal tract without vocal fold vibration, resulting in noise-like spectra. (Bruckert, (2006))

When conducting spectrum analysis of voice, voicing detection typically involves examining the frequency spectrum of the speech signal to determine whether it exhibits characteristics of a voiced or unvoiced sound. This analysis is crucial for understanding various aspects of speech production and can provide valuable information about the phonation type, pitch, and quality of the voice.

In Praat software, "voicing" refers to the process of detecting whether a given segment of speech is voiced or unvoiced. Voiced segments occur when the vocal folds vibrate, producing periodic sound, such as vowels and voiced consonants.

1.9 Jitter- In spectrum analysis of voice, jitter refers to the variation in the fundamental frequency (F0) of the voice signal from one cycle to the next. While jitter is more commonly associated with timing irregularities in the vibration of the vocal folds, in spectrum analysis, it is specifically related to the instability or variability in the pitch of the voice. (González, 2004)

Jitter in spectrum analysis is quantified by measuring the deviation of the fundamental frequency (F0) of the voice signal from its average value over a certain duration of speech. It is typically expressed as a percentage or fraction of the average fundamental frequency. Higher levels of jitter indicate greater instability in the pitch of the voice, whereas lower levels indicate more consistent vocal fold vibration.

In practical terms, jitter in spectrum analysis can be observed as fluctuations or noise in the harmonics of the voice signal. Harmonics are integer multiples of the fundamental frequency and appear as peaks in the frequency spectrum. When there is jitter present, these harmonics may exhibit variations in amplitude or frequency, reflecting the irregularities in vocal fold vibration.

Jitter analysis in spectrum analysis of voice is important for assessing voice quality, diagnosing voice disorders, and monitoring changes in vocal function over time. It is a valuable tool for clinicians and researchers in understanding the underlying mechanisms of voice production and identifying abnormalities or pathologies affecting vocal fold vibration.

1.10 Shimmer- In spectral analysis of voice, shimmer refers to the cycle-to-cycle variation in the amplitude or intensity of the vocal signal. It is a measure of the irregularity or instability in the energy of the voice signal across consecutive cycles of vocal fold vibration. Shimmer is often considered alongside jitter as a measure of voice quality and stability. Measurement: Shimmer is quantified as a percentage or fraction of the average amplitude of the voice signal. Higher levels of shimmer indicate greater variability in vocal fold vibration intensity, while lower levels indicate more consistent vocal fold vibration. Shimmer analysis is valuable for assessing vocal function, diagnosing voice disorders, and monitoring changes in voice quality over time. (Bruckert, (2006))

1.11 Harmonicity- Harmonicity in spectrum analysis of voice refers to the degree to which the frequency components of the voice signal align with the harmonic series. The harmonic series consists of integer multiples of the fundamental frequency (F0) of a sound. In voiced speech sounds, the vibration of the vocal folds produces a fundamental frequency and its harmonics, which are integer multiples of the fundamental frequency.

When a voice signal is analyzed spectrally, harmonicity indicates the extent to which the energy in the signal is concentrated at integer multiples of the fundamental frequency. A highly harmonic voice signal will have strong energy peaks at each harmonic frequency, resulting in a clear and well-defined harmonic structure in the frequency spectrum.

1.11.1 Measurement: Praat calculates harmonicity based on the distribution of energy in the frequency spectrum. A higher Harmonic-to-Noise Ratio (HNR) indicates stronger harmonic structure, while a lower HNR suggests a less well-defined harmonic pattern and potentially more noise in the voice signal.

1.11.2 Interpretation: Harmonicity analysis in Praat provides insights into vocal fold function, voice quality, and the presence of voice disorders. A higher HNR is typically associated with healthy vocal fold vibration and good voice quality, while a lower HNR may indicate vocal fold pathology or voice disorders.

1.12 Correlation of height and weight with voice.

The correlation of height and weight with voice characteristics is an area of interest in fields such as speech science, linguistics, and voice research. While there is some evidence to suggest that height and weight may have correlations with certain voice parameters, such as fundamental frequency (pitch), formant frequencies, and vocal tract length, the relationships are often complex and influenced by various factors. (Lass, (1980))

1.12.1 Fundamental Frequency (Pitch): Research has shown that taller individuals tend to have lower fundamental frequencies (lower pitch) compared to shorter individuals. This relationship is partially attributed to the longer vocal folds typically found in taller individuals. Conversely, weight does not appear to have a significant correlation with fundamental frequency (Lass, (1980))

1.12.2 Formant Frequencies: Formant frequencies, which represent the resonant frequencies of the vocal tract, are influenced by the length and shape of the vocal tract. Taller individuals, with longer vocal tracts, may exhibit lower formant frequencies compared to shorter individuals. However, the relationship between weight and formant frequencies is less clear.

1.12.3 Voice Quality: Height and weight may also have indirect effects on voice quality through their influence on respiratory support, vocal fold tension, and overall vocal health. For example, individuals with larger lung capacities (often correlated with height) may have greater control over airflow and breath support for voice production. (Gonzalez, (2003))

2. REVIEW OF LITERATURE

1- **Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006).** Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 83-89.

This is a study on speaker height estimation from speech, which can assist in voice forensic analysis and automatic speaker identification. The study presents a statistical approach and a fusion method for height estimation, incorporating acoustic models within a non-uniform height bin width Gaussian mixture model structure as well as a formant analysis approach employing linear regression on selected phones. The algorithms achieve competitive performance, significantly decreasing estimation errors compared to past efforts, with a mean average error of 4.89 cm for males and 4.55 cm for females, as reported in the proposed algorithm on TIMIT utterances containing selected phones. The study also explores open set testing using the Multi-session Audio Research Project corpus and publicly available YouTube audio to examine the effect of channel mismatch between training and testing data, providing a realistic open domain testing scenario.

2- **Gonzalez, J. (2003)**. Estimation of speakers' weight and height from speech: A re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and motor skills*, *96*(1), 297-304.

It is a study on speaker height and weight identification from voiced and whispered speech, involving 28 speakers (14 females and 14 males) from West Virginia University. The study aimed to determine the effect of vocal pitch on height and weight identification tasks. The results indicated that listeners were capable of accurately identifying the approximate heights and weights of speakers under both voiced and whispered conditions, with no significant difference in accuracy between the two. This suggests that vocal pitch does not play a major role in height and weight identification tasks. The study also discussed the importance of isolating and defining the acoustic cues in the voice that accurately convey information on speakers' heights and weights, and suggested future research directions in this area. Additionally, the document referenced previous studies on speaker characteristics, including age, sex, race, socioeconomic status, personality, specific identity, and facial features, and their potential value in future investigations. The findings of the study provide empirical evidence on the issue of the importance of vocal pitch in speaker height and weight identification tasks and contribute to the broader understanding of speaker characteristics and perception.

3- **Lass, N. J., & Colt, E. G. (1980)**. A comparative study of the effect of visual and auditory cues on speaker height and weight identification. *Journal of Phonetics*, *8*(3), 277-285.

"A comparative study of the effect of visual and auditory cues on speaker height and weight identification" by Norman J. Lass and Elizabeth G. Colt from West Virginia University compares the impact of visual and auditory cues on speaker height and weight identification. The study involved 30 speakers (15 females and 15 males) and 30 judges who participated in two experimental sessions. In one session, judges made height and weight judgments based on visual cues from photographs, while in the other session, they made judgments based on auditory cues from recorded readings. The results indicated that although differences existed in judges' height and weight estimates between the two experimental conditions, the differences were relatively small, suggesting that the voice alone may convey as much information concerning speakers' heights and weights as visual clues alone. The study also discussed previous research showing that listeners can accurately identify various speaker characteristics from recorded speech samples. The document provides a comprehensive overview of the study's objectives, methodology, findings, and implications, shedding light on the role of visual and auditory cues in speaker height and weight identification.

4- **Van Dommelen, W. A., & Moxness, B. H. (1995)**. Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and speech*, *38*(3), 267-287.

This study investigates the ability of listeners to estimate speaker height and weight based on speech samples. Previous research has suggested that listeners can consistently judge body characteristics from speech, but it's unclear which speech signal parameters they use for these estimations. The study conducted listening tests where male and female participants judged the height and weight of male and female speakers reading isolated words and text paragraphs.The results revealed that both the sex of the speaker and the sex of the listener were significant factors. Male listeners showed significant correlations between estimated and actual height/weight only for male speakers, while female listeners were unable to accurately estimate the height or weight of either male or female speakers.Regression analysis, which involved parameters such as fundamental frequency (F0), formant frequencies, energy below 1 kHz, and speech rate, did not show significant correlations with actual speaker height and weight, except for a correlation between male speaker weight and speech rate. Additionally, the data suggested that listeners correctly used speech rate information to judge male speaker weight, while they incorrectly associated low F0 and formant frequency values with larger speaker body dimensions.

5- Dusan, S. (2005). Estimation of speaker's height and vocal tract length from speech signal. In Interspeech (pp. 1989-1992).

The paper "Estimation of Speaker’s Height and Vocal Tract Length from Speech Signal" by Sorin Dusan explores the relationship between a speaker's height and vocal tract length (VTL) through the analysis of speech signals. Utilizing the TIMIT American English speech corpus, the study investigates correlations between speaker height and various acoustic speech features, including Mel-Frequency Cepstrum Coefficients (MFCC), Linear Predictive Coding (LPC) frequencies, and formant frequencies.The research introduces a novel method for optimal estimation of speaker height and VTL using a phone-based approach, aiming to enhance the accuracy and reliability of height and VTL estimation. Additionally, the paper discusses VTL estimation from speech signals, with a specific focus on the correlation between actual VTL and height values obtained from human subjects.The findings of the study reveal a robust correlation between estimated VTL and height values for the speakers, indicating promising potential for applications in forensic analysis and automatic speech recognition. The paper offers a thorough analysis of various acoustic speech features and their relationships with speaker height, providing valuable insights into the correlation between speaker height and VTL based on speech signal analysis.Overall, the proposed method represents a significant advancement in the field, with implications for forensic analysis and automatic speech recognition systems

6- **Lass, N. J., Hendricks, C. A., & Iturriaga, M. A. (1980).** The consistency of listener judgments in speaker height and weight identification. *Journal of Phonetics*, *8*(4), 439-448.

The purpose of this investigation was to determine the importance of temporal features of speech in speaker height and weight identification. A total of 30 speakers, 15 females and 15 males, recorded a standard prose passage. Three master tapes were constructed from these recordings, representing the three experimental conditions in the study: forward-played, backward-played, and time-compressed tapes. A total of 32 judges participated in three sessions, one for each of the three experimental conditions. In each session they were asked to make direct estimations of the height and weight of each speaker. The order of presentation of the three tapes was counterbalanced. Results indicate that temporal alteration of the speech signal by means of backward-playing adversely affected listener judgments of height and weight. However, time compression of the speech signal did not appear to have a significant effect on listener accuracy. Implications of these findings and suggestions for future research are discussed.

7- **Lass, N. J., Phillips, J. K., & Bruchey, C. A. (1980**). The effect of filtered speech on speaker height and weight identification. *Journal of Phonetics*, *8*(1), 91-100.

The study aimed to assess the significance of different segments of the speech spectrum in determining speaker height and weight. It discovered that listeners can accurately estimate speakers' heights and weights based on recorded speech samples. Additionally, the study investigated the impact of filtering the speech signal on height and weight estimation and concluded that filtering had no substantial effect on accuracy.Furthermore, the results indicated that both vocal tract resonance characteristics and laryngeal fundamental frequency contribute equally to height and weight estimation tasks. The study recommended further research to isolate specific acoustic cues that contribute to precise identification and to explore narrower regions of the speech spectrum.Moreover, the study underscored the potential for future theoretical and applied investigations in various domains. Overall, the findings offered valuable insights into the acoustic cues employed by listeners in speaker height and weight estimation and emphasized the necessity for ongoing research to improve understanding and potential applications across diverse fields.

9- **Lass, N. J., Dicola, G. A., Beverly, A. S., Barbera, C., Henry, K. G., & Badali, M. K. (1979)**. The effect of phonetic complexity on speaker height and weight identification. *Language and Speech*, *22*(4), 297-309.

The purpose was to determine the effect of phonetic complexity on speaker height and weight identification in an attempt to establish the minimal acoustic cues necessary for distinguishing speakers' heights and weights. A total of 28 speakers, 14 females and 14 males, recorded 16 utterances representing four levels of phonetic complexity: isolated vowels, monosyllabic words, bisyllabic words and sentences. Eight master tapes were constructed, one for each of the four levels of phonetic complexity and the two sex groups. Twenty-four female judges participated in four experimental sessions, one for each level of phonetic complexity. In each session they heard the female and male tapes of one kind of stimulus and were asked to make direct estimations of the speakers' heights and weights. Results show that listeners are capable of accurate height and weight identification at all levels of phonetic complexity investigated. Differences in listener accuracy are, on average, only 0.41 inches and 1.44 lbs. Moreover, no regular, progressive trend is evident in listener accuracy from the simplest to the most complex stimuli. Implications of these findings and suggestions for future research are discussed.

10- **Künzel, H. J. (1989)**. How well does average fundamental frequency correlate with speaker height and weight?. *Phonetica*, *46*(1-3), 117-125.

The purpose of this investigation is to shed some more light on the conflicting views about a number of acoustic parameters which might carry information on some general somatic features of the speaker. In an experiment, average fundamental frequency (F_0) values of 105 male and 78 female adult subjects were correlated with their individual height and weight data. No significant correlations between acoustic and physical parameters were found. The results are discussed with respect to earlier studies with completely different approaches to the issue, namely direct estimation of physical traits from the speech signal by listeners.

11- **Jaid, U. H., & Abdulhassan, A. K. (2023)**. Fuzzy-Based Ensemble Feature Selection for Automated Estimation of Speaker Height and Age Using Vocal Characteristics. *IEEE Access*.

This work introduces a novel method for automatic speaker height and age estimation by employing fuzzy-based ensemble feature selection with speech parameters. The approach begins by deriving initial feature importance from various feature selection (FS) methods. These importance values are then organized into a matrix, converted to ranks, and inputted into the fuzzy c-means (FCM) algorithm to generate final feature rankings and identify the most distinctive features for estimation.Notably, the proposed approach offers flexibility in combining results from multiple feature importance methods into an ensemble approach or selecting the best features based on importance from a single method, without necessitating a predetermined number of top features. The effectiveness of this method was evaluated through experiments conducted on acoustic features extracted using the OpenSMILE toolkit from the TIMIT dataset.The results demonstrate the efficacy of the proposed approach in selecting the most informative features, surpassing similar studies on the same dataset. Particularly, promising results were obtained with mean absolute error (MAE) values of 5.4 and 4.71 for height estimation, and 5.38 and 5.24 for age estimation among males and females, respectively.

12- **Poorjam, A. H., Bahari, M. H., & Van Hamme, H. (2015, March)**. Speaker weight estimation from speech signals using a fusion of the i-vector and NFA frameworks. In *2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP)* (pp. 118-123). IEEE.

This method utilizes two frameworks, namely the i-vector framework and the Non-negative Factor Analysis (NFA) framework, to model each utterance. The i-vector framework is based on factor analysis on Gaussian Mixture Model (GMM) mean supervectors, while the NFA framework employs constrained factor analysis on GMM weights. These frameworks leverage information from Gaussian means and weights, respectively.Subsequently, a feature-level fusion of i-vectors and NFA vectors is performed to exploit available information from both frameworks. Finally, a least-squares support vector regression (LS-SVR) technique is applied to estimate the weight of speakers from the given utterances.The proposed approach is evaluated on telephone speech signals from the National Institute of Standards and Technology (NIST) 2008 and 2010 Speaker Recognition Evaluation (SRE) corpora. Experimental results conducted over 2339 utterances demonstrate promising results. The correlation coefficients between actual and estimated weights of male and female speakers are reported as 0.56 and 0.49, respectively. These results indicate the effectiveness of the proposed method in speaker weight estimation.

13- **Arsikere, H., Leung, G. K., Lulich, S. M., & Alwan, A. (2013)**. Automatic estimation of the first three subglottal resonances from adults' speech signals with application to speaker height estimation. *Speech Communication*, *55*(1), 51-70.

This paper proposes a novel algorithm for estimating the first three SGRs (Sg1,Sg2 and Sg3) from *continuous* adults' speech. While Sg1 and Sg2 are estimated based on the phonological distinction they provide between vowel categories, Sg3 is estimated based on its correlation with Sg2. The RMS estimation errors (approximately 30, 60 and 100 Hz for Sg1,Sg2 and Sg3, respectively) are not only comparable to the standard deviations in the measurements, but also are independent of vowel content and language (English and Spanish). Since SGRs correlate with speaker height while remaining roughly constant for a given speaker (unlike vocal tract parameters), the proposed algorithm is applied to the task of height estimation using speech signals. The proposed height estimation method matches state-of-the-art algorithms in performance (mean absolute error = 5.3 cm), but uses much less training data and a much smaller feature set. Our results, with additional analysis of physiological data, suggest the existence of a limit to the accuracy of speech-based height estimation.

14- **González, J. (2004)**. Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of phonetics*, *32*(2), 277-287.

This paper explores the relationship between formant frequencies and body size in human adults through two experiments.In Experiment I, correlation coefficients were calculated between acoustic features of five Spanish vowels produced by 82 speakers and the speakers' heights and weights. In Experiment II, correlations were determined from formant parameters derived from long-term average analysis of connected speech from 91 speakers.Contrary to findings in macaque vocalizations by Fitch (1997), the study reveals that the relationship between formant parameters and body size within sexes is notably weak in human adults. However, correlations within the female group were found to be stronger than those within the male group.The results suggest that individual vocal tract development in humans is relatively unaffected by skeletal size constraints, likely due to the descent of the larynx from its standard mammal position. This disassociation between vocal tract and body size appears to be more pronounced in human males.

3. Methodology

Sample population size- A total of 150 children, aged between 7 to 18 years, were selected for this study. This cohort was further subdivided into three groups: 50 participants from high school, 50 from middle school, and 50 from primary school.

Data Collection Instruments-Voice samples were recorded from each participant using high-clarity headphones equipped with a speaker to ensure optimal audio quality. Additionally, the weight of each individual was measured using a weighing machine, while their height was measured using either a stadiometer or a measuring tape. The age of each speaker was also recorded.

Spectrum Analysis- To analyze the voice samples, spectrum analysis was conducted using the Praat software. This allowed for the examination of various vocal parameters such as pitch, pulse, voicing, jitter, shimmer and harmonicity

Speech Sentence: All participants were instructed to recite the sentence: "The quick brown fox jumps over the lazy dog." This standardized sentence was chosen to ensure consistency across all samples.

4.Data Analysis

Following the collection of voice samples and relevant measurements from the participants, the average range of age, height, and weight within each subgroup was calculated. This allowed for a comprehensive understanding of the demographic characteristics of the sample population. (Lass N. J., 1979). Following the collection of voice samples using Praat software, a comprehensive data analysis was conducted to examine various vocal parameters. The analysis included the following steps:

1.Preprocessing: Before analysis, the voice samples were preprocessed to remove any background noise or artifacts that could affect the accuracy of the results. This step ensured that only clear and relevant data were included in the analysis.
2.Frequency Analysis: The frequency of each voice sample was analyzed to determine the fundamental frequency (F0) or pitch. This analysis provided insights into the natural pitch range of each participant's voice. (5. Van Dommelen, (1995))
3.Spectrum Analysis: Spectral analysis was conducted to examine the distribution of energy across different frequency bands in the voice samples.
4.Pitch Contour Analysis: The pitch contour of each voice sample was plotted and analyzed to observe variations in pitch over time. This analysis provided insights into the prosodic features of the speech, such as intonation and stress patterns.
5.Pulse Analysis: Pulse analysis was performed to extract the fundamental frequency (F0) contour from each voice sample. Praat's pulse analysis functions were used to measure the periodicity of vocal fold vibration, providing information about vocal stability and regularity.
6.Voicing Analysis: Voicing analysis was conducted to differentiate between voiced and unvoiced segments within the voice signal. Praat's voicing detection algorithms were applied to classify time frames as voiced or unvoiced based on amplitude and periodicity criteria.
7.Jitter Analysis: Jitter analysis was performed to quantify the cycle-to-cycle variations in F0, indicating the stability of vocal fold vibration. Praat calculated jitter measures, such as average absolute jitter (AJ) and relative average jitter (RAP), to assess vocal irregularity and instability.
8. Shimmer Analysis: Shimmer analysis was conducted to measure the cycle-to-cycle variations in voice intensity, reflecting vocal stability and consistency. Praat calculated shimmer measures, including average absolute shimmer (AS) and relative average shimmer (RAP), to assess vocal variability and roughness.
9.Harmonicity Analysis: Harmonicity analysis was carried out to evaluate the degree of periodicity and harmonic structure in the voice signal. Praat calculated harmonicity measures, such as Harmonics-to-Noise Ratio (HNR) and cepstral peak prominence (CPP), to assess vocal clarity and purity. (Dusan, (2005))

Finally, the results of the data analysis were interpreted to draw conclusions about the variations in voice parameters among children aged 12 to 18 years. Any correlations or relationships between vocal characteristics and demographic factors such as height, weight were also examined and discussed.

Table 1 – Height 4-4.5ft, voice parameters readings.

HEIGHT	PITCH	PULSES	VOICING	JITTER	SHIMMER	HARMONICITY
4-4.5	332.97HZ	903	11	3.46%	17.00%	0.764396
4-4.5	332.78HZ	799	6	3.46%	16.52%	0.765731
4-4.5	327.98HZ	635	11	3.45%	17.53%	0.75955
4-4.5	299.98HZ	342	9	3.53%	20.43%	0.717346
4-4.5	331.78HZ	1269	36	3.07%	19.04%	0.707627
4-4.5	302.67HZ	507	10	2.05%	14.20%	0.817383
4-4.5	302.67HZ	1300	19	3.90%	18.59%	0.740128
4-4.5	302.67HZ	1054	22	2.09%	16.71%	0.765893
4-4.5	302.67HZ	385	8	3.85%	20.67%	0.645879
4-4.5	302.67HZ	829	18	2.53%	13.17%	0.826823
4-4.5	332.98HZ	776	22	1.40%	14.45%	0.794643
4-4.5	322.76HZ	749	30	3.28%	16.64%	0.766805
4-4.5	334.96HZ	411	17	2.56%	17.85%	0.72215
4-4.5	327.76HZ	885	9	3.53%	17.91%	0.68751
4-4.5	328.96HZ	1314	10	2.44%	2.44%	0.82337
4-4.5	322.56HZ	567	12	3.76%	17.58%	0.746085
4-4.5	298.94HZ	501	18	3.40%	17.28%	0.78201

Graph 2- Height 4-4.5ft, voice parameters variations

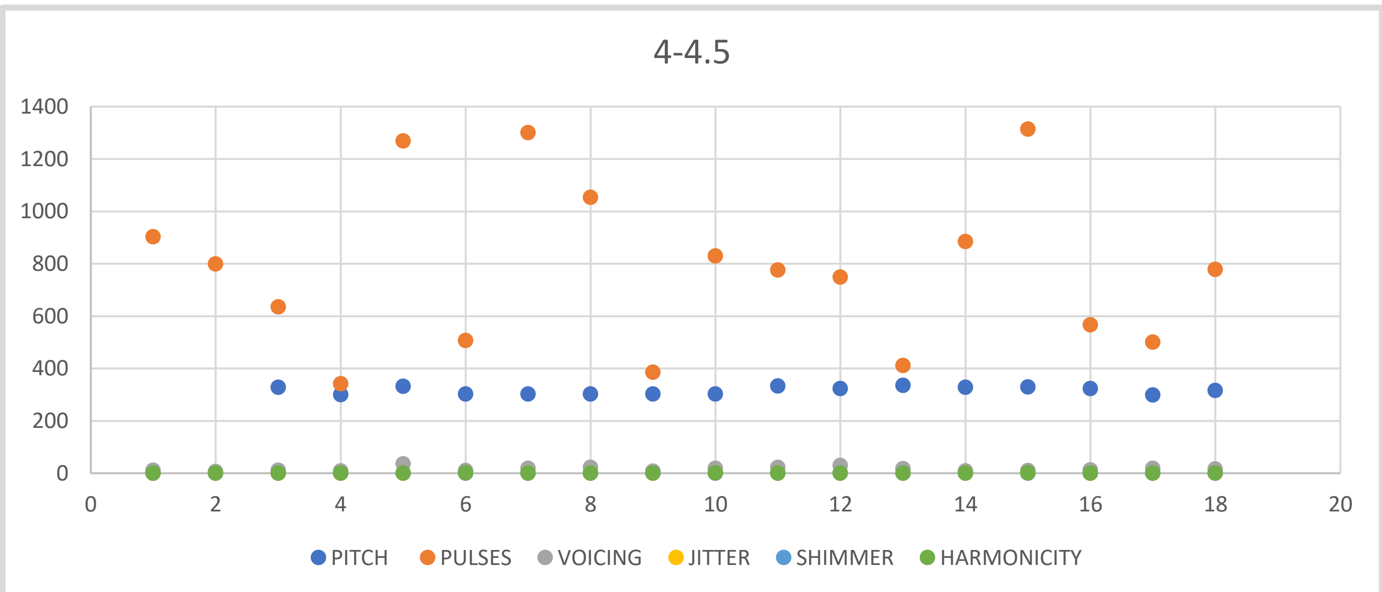


Table 2- Height 4.6-5ft, voice parameters readings.

HEIGHT	PITCH	PULSES	VOICING	JITTER	SHIMMER	HARMONICITY
4.6-5	268.252 Hz	524	12	3.65%	17.81%	0.742667
4.6-5	270.844 Hz	555	12	3.74%	13.74%	0.405292
4.6-5	281.764HZ	1020	32	3.07%	17.89%	0.746085
4.6-5	293.255HZ	879	18	3.24%	17.89%	0.746085
4.6-5	254.666HZ	990	22	3.24%	17.89%	0.746085
4.6-5	301.455HZ	658	15	3.24%	17.89%	0.746085
4.6-5	278.901HZ	1256	12	3.24%	17.89%	0.746085
4.6-5	275.998HZ	1022	33	3.24%	17.89%	0.746085
4.6-5	282.667HZ	1033	26	3.24%	17.89%	0.746085
4.6-5	268.236HZ	1044	27	3.24%	17.89%	0.746085
4.6-5	290.456HZ	1055	38	3.24%	17.89%	0.746085
4.6-5	301.226HZ	1066	25	3.24%	17.89%	0.746085
4.6-5	239.465HZ	1095	35	3.24%	17.89%	0.746085
4.6-5	262.425HZ	1037	27	3.24%	17.89%	0.746085
4.6-5	298.789HZ	1216	24	3.24%	17.89%	0.746085
4.6-5	298.789HZ	1235	13	3.24%	17.89%	0.746085
4.6-5	298.789HZ	1583	20	3.24%	17.89%	0.746085
4.6-5	298.789HZ	1267	23	3.24%	17.89%	0.746085
4.6-5	298.789HZ	1257	24	2.89%	17.89%	0.746085
4.6-5	256.809HZ	1904	17	3.89%	17.89%	0.707627
4.6-5	290.456HZ	1278	25	2.67%	17.89%	0.707627
4.6-5	290.456HZ	1456	15	2.44%	17.89%	0.707627
4.6-5	290.456HZ	879	16	2.44%	17.89%	0.707627
4.6-5	290.456HZ	798	16	2.44%	17.89%	0.707627
4.6-5	298.560HZ	958	18	2.44%	17.89%	0.707627
4.6-5	290.456HZ	989	29	2.44%	17.89%	0.707627
4.6-5	290.456HZ	900	33	2.44%	17.89%	0.707627
4.6-5	290.456HZ	247	30	2.44%	17.89%	0.707627
4.6-5	310.289HZ	1057	32	2.44%	17.89%	0.707627
4.6-5	310.289HZ	1094	45	2.44%	17.89%	0.707627
4.6-5	310.289HZ	1045	37	2.44%	29.56%	0.707627
4.6-5	310.289HZ	1267	21	2.44%	23.45%	0.746085
4.6-5	310.289HZ	1287	18	2.44%	21.84%	0.746085
4.6-5	310.289HZ	1589	19	2.44%	21.84%	0.746085
4.6-5	310.289HZ	1345	10	2.44%	21.84%	0.746085
4.6-5	310.289HZ	1278	15	2.44%	21.84%	0.746085
4.6-5	310.289HZ	1248	41	2.44%	21.84%	0.746085
4.6-5	310.289HZ	1290	14	2.44%	21.84%	0.746085
4.6-5	310.289HZ	1359	19	3.46%	21.84%	0.746085
4.6-5	310.289HZ	890	23	3.68%	21.84%	0.746085
4.6-5	310.289HZ	955	26	3.09%	21.84%	0.746085
4.6-5	310.289HZ	945	27	3.90%	21.84%	0.746085

4.6-5	310.289HZ	993	33	4.90%	21.84%	0.746085
4.6-5	310.289HZ	895	89	3.97%	21.84%	0.746085
4.6-5	310.289HZ	1902	45	3.97%	21.84%	0.746085
4.6-5	310.289HZ	1946	34	3.97%	21.84%	0.746085
4.6-5	310.289HZ	1930	38	3.97%	23.67%	0.736588
4.6-5	310.289HZ	1256	17	3.97%	25.78%	0.762284
4.6-5	310.289HZ	1349	19	3.97%	24.76%	0.762789
4.6-5	234.765HZ	1468	16	3.97%	29.92%	0.702678
4.6-5	255.897HZ	1368	20	3.97%	16.90%	0.789032
4.6-5	345.895HZ	1287	21	3.97%	15.89%	0.789032
4.6-5	345.895HZ	1368	24	3.97%	17.45%	0.789032
4.6-5	345.895HZ	1245	25	3.97%	17.90%	0.789032
4.6-5	345.895HZ	1345	18	3.97%	20.67%	0.789032
4.6-5	345.895HZ	1542	31	3.97%	19.56%	0.789032
4.6-5	345.895HZ	1324	30	3.97%	21.23%	0.765398
4.6-5	345.895HZ	1245	22	3.97%	17.45%	0.765348
4.6-5	345.895HZ	1267	17	3.97%	26.97%	0.765322
4.6-5	345.895HZ	1363	19	3.97%	14.58%	0.245839

Graph 2- Height 4.6-5ft, voice parameters variations

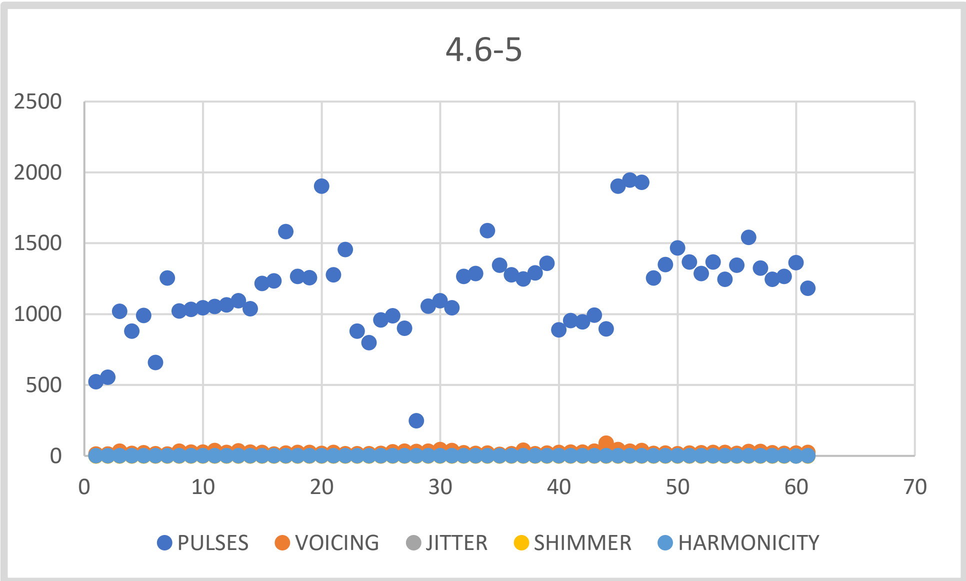


Table 3- Height 5.1-5.5ft, voice parameters reading.

HEIGHT	PITCH	PULSES	VOICING	JITTER	SHIMMER	HARMONICITY
5.1-5	327.515hz	1308	12	2.42%	13.35%	0.823768
5.1-5	327.515Hz	1196	18	2.99%	14.41%	0.788512
5.1-5	327.515Hz	1196	18	2.42%	14.41%	0.788512
5.1-5	365.876hz	989	9	3.23%	19.43%	0.752303
5.1-5	265.987hz	514	12	3.21%	15.82%	0.756775
5.1-5	332.978hz	683	6	3.23%	14.49%	0.808762
5.1-5	3675.857z	579	8	2.94%	16.44%	0.809022
5.1-5	327.515Hz	575	7	2.52%	16.63%	0.822467
5.1-5	327.515Hz	417	5	2.87%	14.85%	0.797083
5.1-5	327.515Hz	302	14	3.46%	20.90%	0.661139
5.1-5	453.786hz	589	18	2.95%	14.78%	0.789688
5.1-5	327.515Hz	478	19	3.78%	17.14%	0.788744
5.1-5	327.515Hz	487	20	3.89%	14.82%	0.829829
5.1-5	327.515Hz	397	25	3.98%	14.83%	0.723145
5.1-5	356.987hz	578	17	4.89%	16.98%	0.774763
5.1-5	312.578hz	978	12	3.86%	16.98%	0.774763
5.1-5	327.515Hz	657	23	3.86%	16.98%	0.774763
5.1-5	424.986hz	564	12	3.86%	16.98%	0.774763
5.1-5	322.767hz	436	23	3.86%	16.98%	0.774763
5.1-5	324.767hz	435	12	3.86%	16.98%	0.774763
5.1-5	356.265hz	356	9	3.86%	16.98%	0.774763
5.1-5	365.796hz	367	19	3.86%	14.78%	0.774763
5.1-5	365.977hz	389	28	3.86%	15.78%	0.774763
5.1-5	265.876hz	478	38	3.65%	16.56%	0.774763
5.1-5	327.515Hz	598	12	4.87%	16.56%	0.774763
5.1-5	345.897hz	598	18	3.76%	16.56%	0.774763
5.1-5	435.767hz	987	11	3.76%	16.56%	0.775643
5.1-5	327.515Hz	1223	23	3.76%	16.56%	0.775635
5.1-5	378.214hz	1020	22	3.76%	16.56%	0.775456
5.1-5	389.231hz	455	12	3.76%	17.47%	0.774354
5.1-5	327.515Hz	476	18	3.76%	16.98%	0.775434
5.1-5	327.515Hz	458	29	4.76%	16.98%	0.775644
5.1-5	876.765hz	456	16	3.67%	16.98%	0.776543
5.1-5	863.965hz	325	19	4.87%	16.98%	0.756437
5.1-5	765.877hz	677	14	3.39%	16.98%	0.786954
5.1-5	410.786hz	987	17	3.35%	16.98%	0.786543
5.1-5	311.789hz	876	18	4.67%	16.75%	0.776545
5.1-5	346.964hz	768	12	4.78%	16.98%	0.776754
5.1-5	344.788hz	567	23	3.89%	16.98%	0.776543

Graph 3- Height 5.1-5.5ft, voice parameters variations.

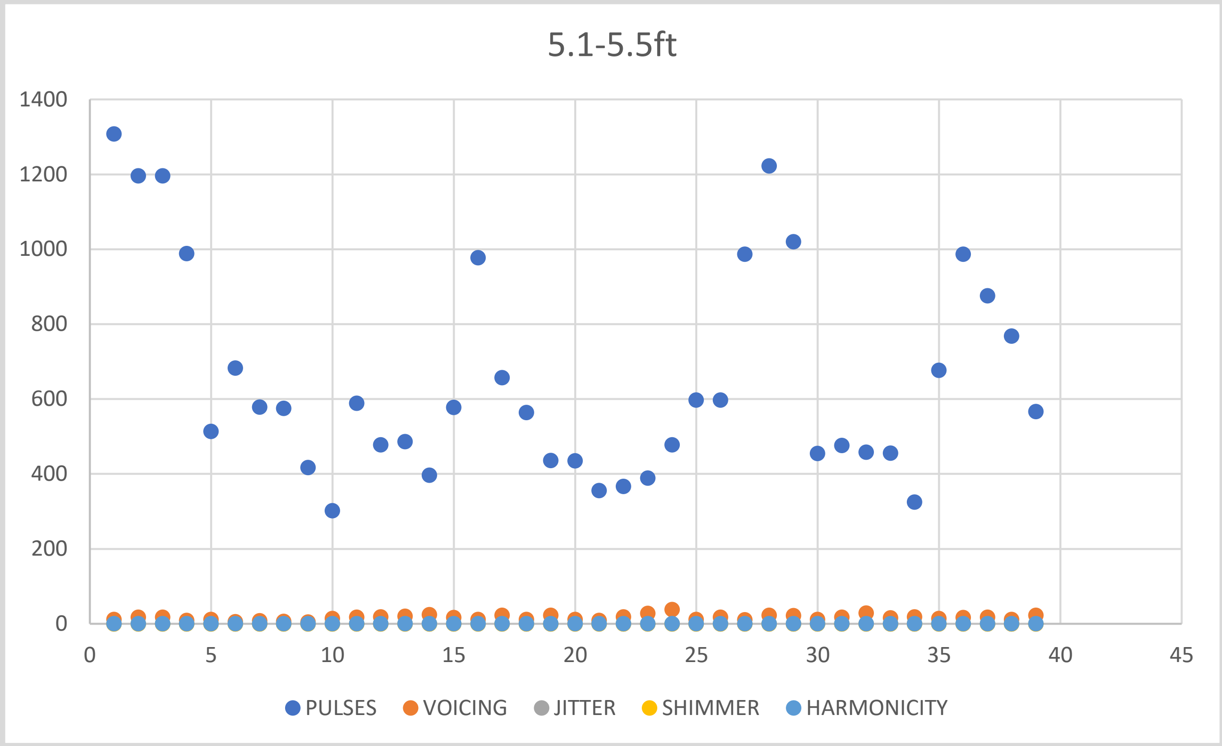
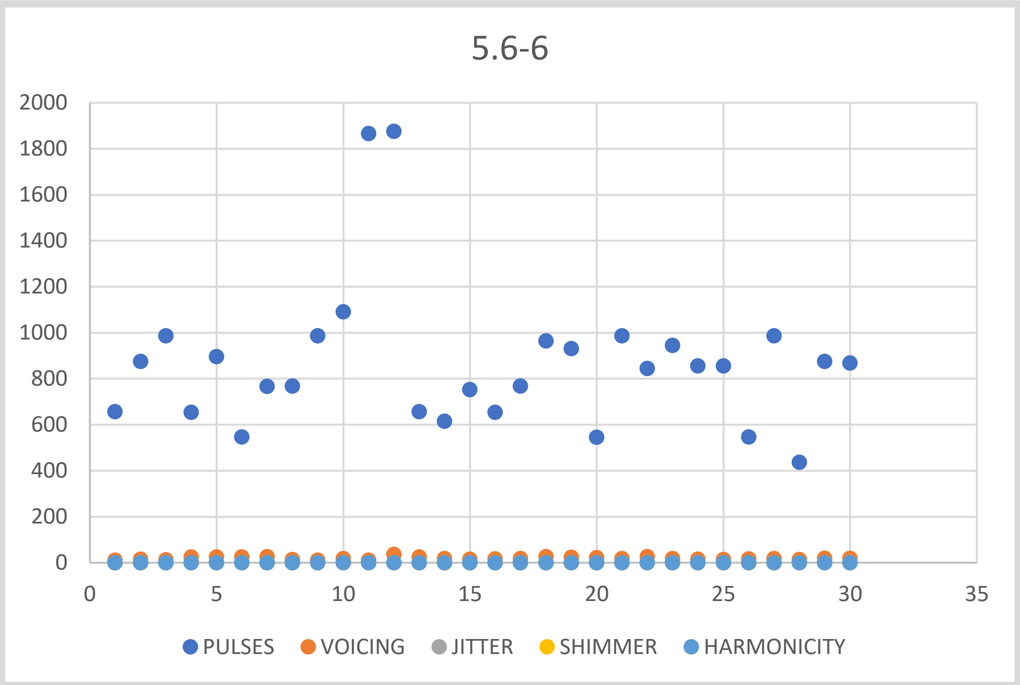


Table 4- Height 5.5-6ft, voice parameters readings.

HEIGHT	PITCH	PULSES	VOICING	JITTER	SHIMMER	HARMONICITY
5.6-6	345.876HZ	657	12	4.88%	17.87%	0.776547
5.6-6	387.987HZ	876	15	4.97%	17.87%	0.745333
5.6-6	357.954HZ	987	13	4.97%	17.87%	0.775643
5.6-6	345.098HZ	654	25	4.99%	17.87%	0.756447
5.6-6	346.098HZ	897	25	4.99%	17.87%	0.756447
5.6-6	456.987HZ	548	25	4.99%	17.87%	0.756447
5.6-6	234.987HZ	767	27	4.99%	17.87%	0.756447
5.6-6	367.9807HZ	769	14	4.99%	17.87%	0.756447
5.6-6	325.987HZ	987	11	4.99%	17.87%	0.756447
5.6-6	357.879HZ	1091	19	4.99%	17.87%	0.756447
5.6-6	456.096HZ	1867	11	4.99%	17.87%	0.756447
5.6-6	5436.987HZ	1876	37	4.85%	17.87%	0.776435
5.6-6	643.789HZ	657	26	4.65%	17.87%	0.776643
5.6-6	376.678HZ	616	18	4.87%	17.87%	0.779874
5.6-6	367.988HZ	754	16	4.34%	16.97%	0.776542
5.6-6	376.986HZ	654	17	4.77%	17.34%	0.787456
5.6-6	376.877HZ	769	18	4.98%	16.97%	0.784535
5.6-6	265.878HZ	965	27	4.34%	16.45%	0.785456
5.6-6	376.567HZ	932	24	4.54%	16.45%	0.785454
5.6-6	398.597HZ	546	22	4.43%	16.45%	0.798642
5.6-6	365.857HZ	987	18	4.98%	16.45%	0.798765
5.6-6	367.987HZ	845	27	4.76%	16.45%	0.879754
5.6-6	387.765HZ	945	19	4.65%	16.45%	0.798453

5.6-6	365.097HZ	856	16	4.98%	17.98%	0.775428
5.6-6	308.985HZ	856	14	3.98%	17.76%	0.775644
5.6-6	301.976HZ	547	17	3.87%	17.87%	0.785322
5.6-6	309.865HZ	987	18	3.98%	16.76%	0.770546
5.6-6	376.986HZ	437	14	3.87%	16.67%	0.772347
5.6-6	398.097HZ	876	20	3.96%	17.87%	0.771232

Graph 4-Height 5.5-6ft, voice parameters variations



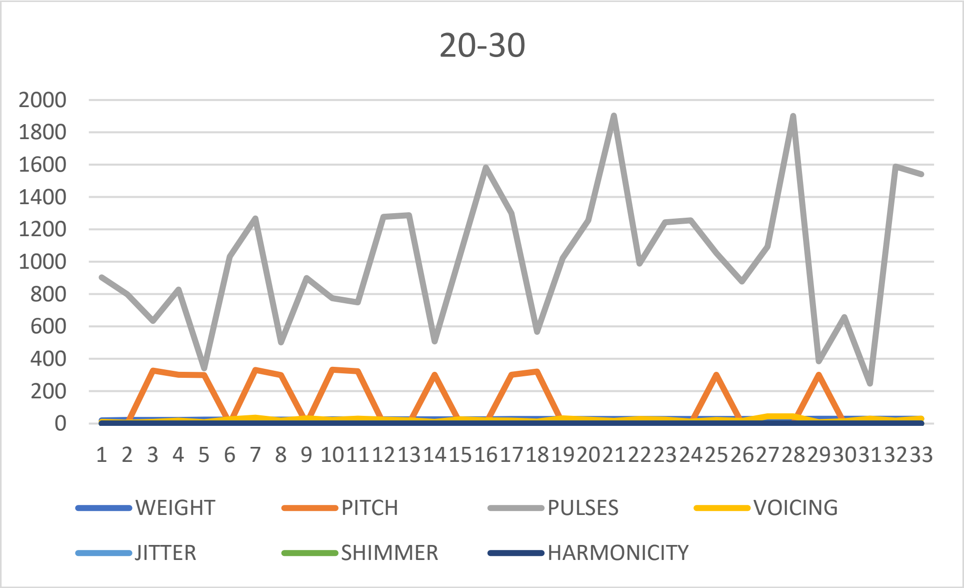
From the above data analysis of different height ranges that is 4-4.5ft, 4.6-5ft, 5.1-5ft, 5.1-5.5ft, 5.5-6ft the voice parameters reading from the spectrographical analysis showed high variations in pulse and pitch, and moderate variations in the voicing, jitter, shimmer, harmonicity.

Table 5- Weight 20-30kg, voice parameters reading

WEIGHT	PITCH	PULSES	VOICING	JITTER	SHIMMER	HARMONICITY
20	332.97HZ	903	11	3.46%	17.00%	0.764396
22	332.78I	799	6	3.46%	16.52%	0.765731
23	327.98	635	11	3.45%	17.53%	0.75955
23	302.67	829	18	2.53%	13.17%	0.826823
24	299.98	342	9	3.53%	20.43%	0.717346
24	282.667HZ	1033	26	3.24%	17.89%	0.746085
25	331.78	1269	36	3.07%	19.04%	0.707627
25	298.9	501	18	3.40%	17.28%	0.78201
25	290.456HZ	900	33	2.44%	17.89%	0.707627
26	332.98	776	22	1.40%	14.45%	0.794643

26	322.76	749	30	3.28%	16.64%	0.766805
26	290.456HZ	1278	25	2.67%	17.89%	0.707627
26	345.895HZ	1287	21	3.97%	15.89%	0.789032
27	302.67	507	10	2.05%	14.20%	0.817383
27	268.236HZ	1044	27	3.24%	17.89%	0.746085
27	298.789HZ	1583	20	3.24%	17.89%	0.746085
28	302.67	1300	19	3.90%	18.59%	0.740128
28	322.56	567	12	3.76%	17.58%	0.746085
28	275.998HZ	1022	33	3.24%	17.89%	0.746085
28	298.789HZ	1257	24	2.89%	17.89%	0.746085
28	256.809HZ	1904	17	3.89%	17.89%	0.707627
28	290.456HZ	989	29	2.44%	17.89%	0.707627
28	345.895HZ	1245	25	3.97%	17.90%	0.789032
29	278.901HZ	1256	12	3.24%	17.89%	0.746085
29	302.67	1054	22	2.09%	16.71%	0.765893
29	290.456HZ	879	16	2.44%	17.89%	0.707627
29	310.289HZ	1094	45	2.44%	17.89%	0.707627
29	310.289HZ	1902	45	3.97%	21.84%	0.746085
30	302.67	385	8	3.85%	20.67%	0.645879
30	301.455HZ	658	15	3.24%	17.89%	0.746085
30	290.456HZ	247	30	2.44%	17.89%	0.707627
30	310.289HZ	1589	19	2.44%	21.84%	0.746085
30	345.895HZ	1542	31	3.97%	19.56%	0.789032

Graph 5,6-Weight 20-30, voice parameters variations.



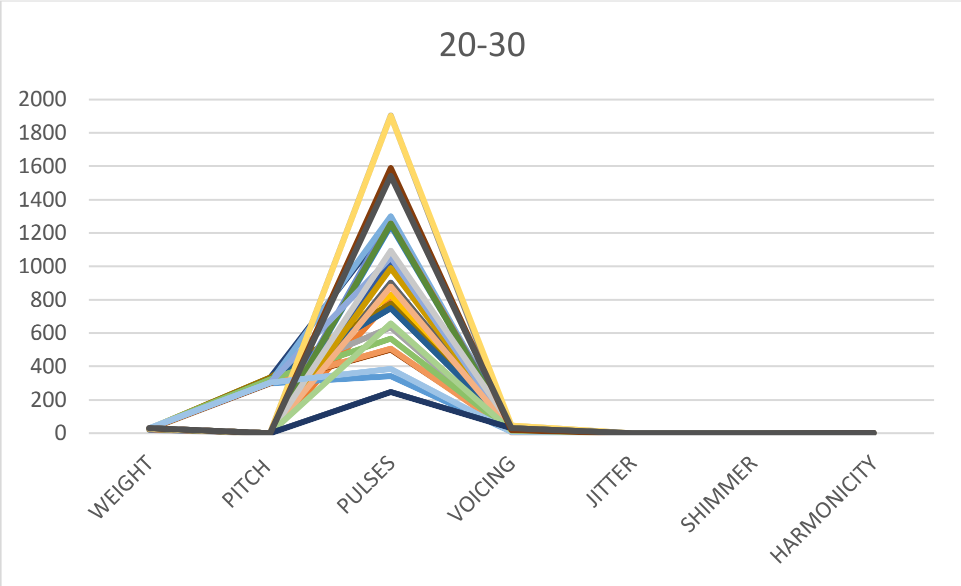
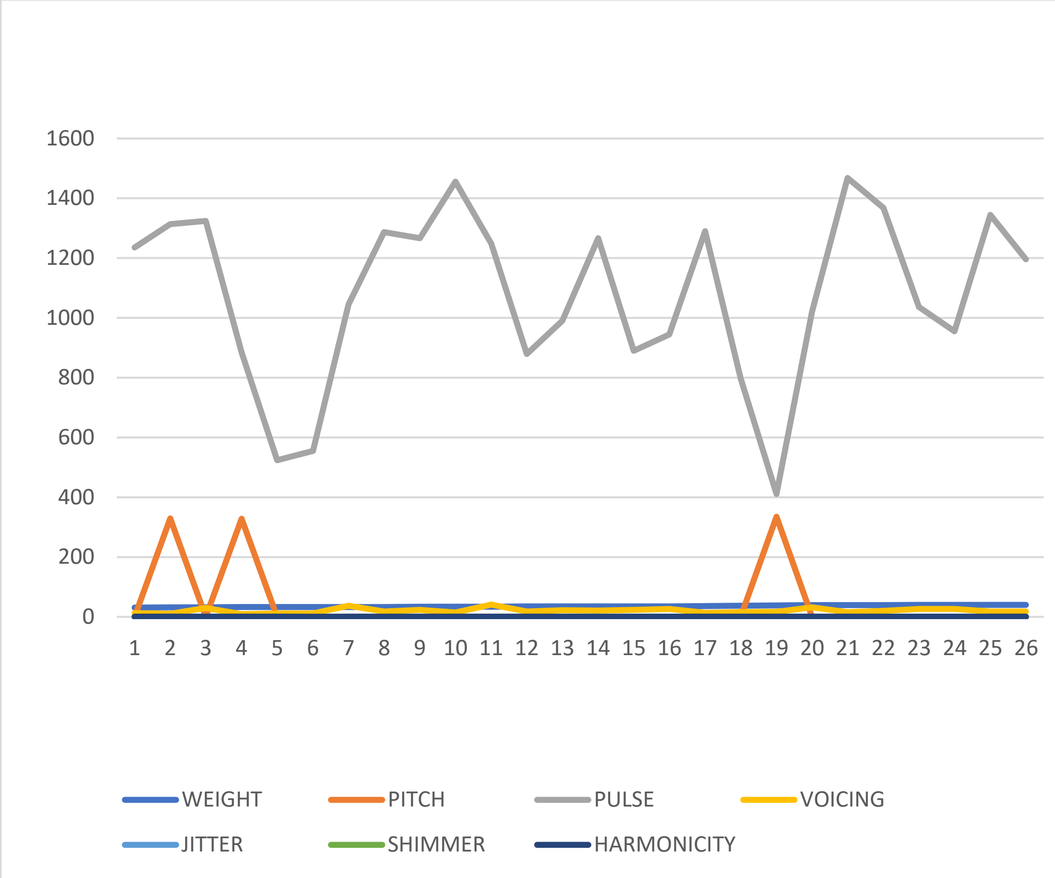


Table 6- Weight 31-40, voice parameters reading.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
31	298.789HZ	1235	13	3.24%	17.89%	0.746085
32	328.964HZ	1314	10	2.44%	2.44%	0.82337
32	345.895HZ	1324	30	3.97%	21.23%	0.765398
33	327.765HZ	885	9	3.53%	17.91%	0.68751
33	268.252Hz	524	12	3.65%	17.81%	0.742667
33	270.844Hz	555	12	3.74%	13.74%	0.405292
33	310.289HZ	1045	37	2.44%	29.56%	0.707627
33	310.289HZ	1287	18	2.44%	21.84%	0.746085
34	298.789HZ	1267	23	3.24%	17.89%	0.746085
34	290.456HZ	1456	15	2.44%	17.89%	0.707627
34	310.289HZ	1248	41	2.44%	21.84%	0.746085
35	293.255HZ	879	18	3.24%	17.89%	0.746085
35	254.666HZ	990	22	3.24%	17.89%	0.746085
35	310.289HZ	1267	21	2.44%	23.45%	0.746085
35	310.289HZ	890	23	3.68%	21.84%	0.746085
35	310.289HZ	945	27	3.90%	21.84%	0.746085
36	310.289HZ	1290	14	2.44%	21.84%	0.746085
37	290.456HZ	798	16	2.44%	17.89%	0.707627
38	334.968HZ	411	17	2.56%	17.85%	0.72215
39	281.764HZ	1020	32	3.07%	17.89%	0.746085
39	234.765HZ	1468	16	3.97%	29.92%	0.702678
39	255.897HZ	1368	20	3.97%	16.90%	0.789032
40	262.425HZ	1037	27	3.24%	17.89%	0.746085
40	310.289HZ	955	26	3.09%	21.84%	0.746085
40	345.895HZ	1345	18	3.97%	20.67%	0.789032
40	327.515 Hz	1196	18	2.99%	14.41%	0.788512

Graph 7,8- Weight 31-40



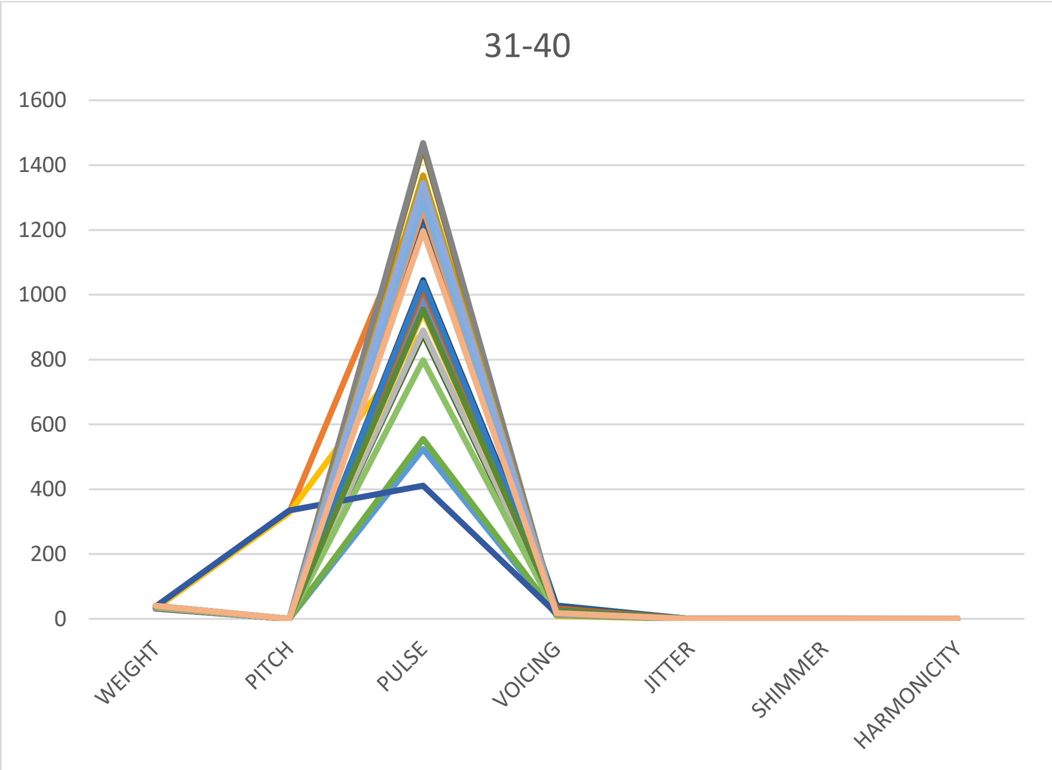
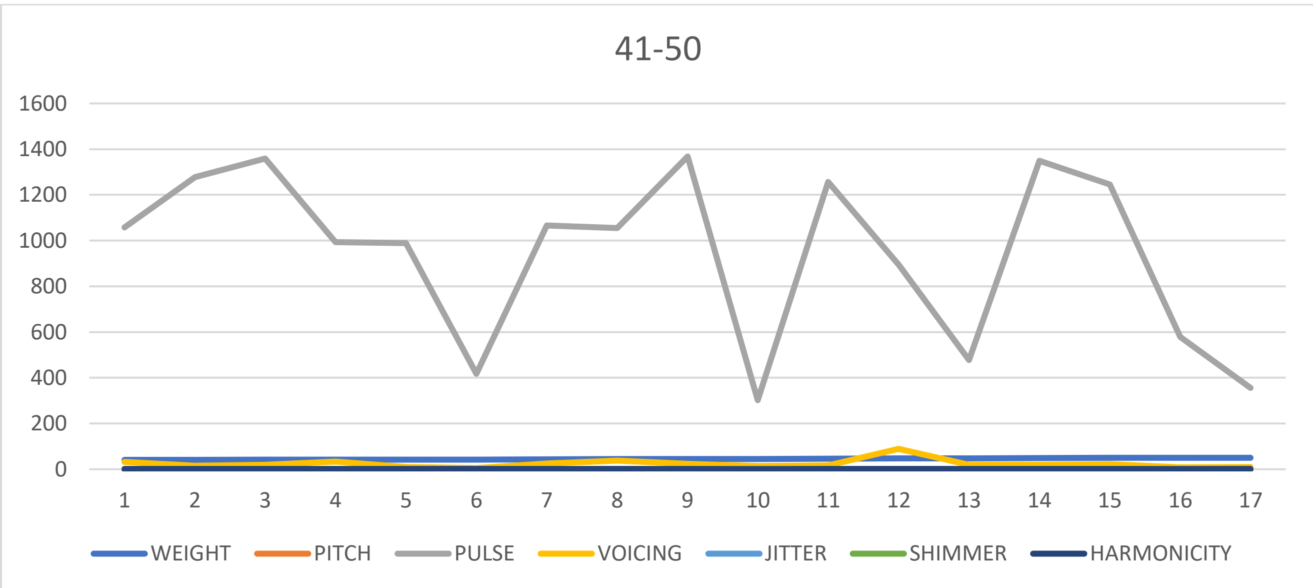


Table 7- Weight 41-50kgs, voice parameters readings.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
41	310.289HZ	1057	32	2.44%	17.89%	0.707627
41	310.289HZ	1278	15	2.44%	21.84%	0.746085
42	310.289HZ	1359	19	3.46%	21.84%	0.746085
42	310.289HZ	993	33	4.90%	21.84%	0.746085
42	327.515Hz	989	9	3.23%	19.43%	0.752303
42	327.515Hz	417	5	2.87%	14.85%	0.797083
43	301.226HZ	1066	25	3.24%	17.89%	0.746085
44	290.456HZ	1055	38	3.24%	17.89%	0.746085
44	345.895HZ	1368	24	3.97%	17.45%	0.789032
45	327.515Hz	302	14	3.46%	20.90%	0.661139
46	310.289HZ	1256	17	3.97%	25.78%	0.762284
47	310.289HZ	895	89	3.97%	21.84%	0.746085
48	327.515Hz	478	19	3.78%	17.14%	0.788744
49	310.289HZ	1349	19	3.97%	24.76%	0.762789
50	345.895HZ	1245	22	3.97%	17.45%	0.765348
50	327.515Hz	579	8	2.94%	16.44%	0.809022
50	327.515Hz	356	9	3.86%	16.98%	0.774763

Graph 9,10- Weight 41-50kgs, voice parameters variations.



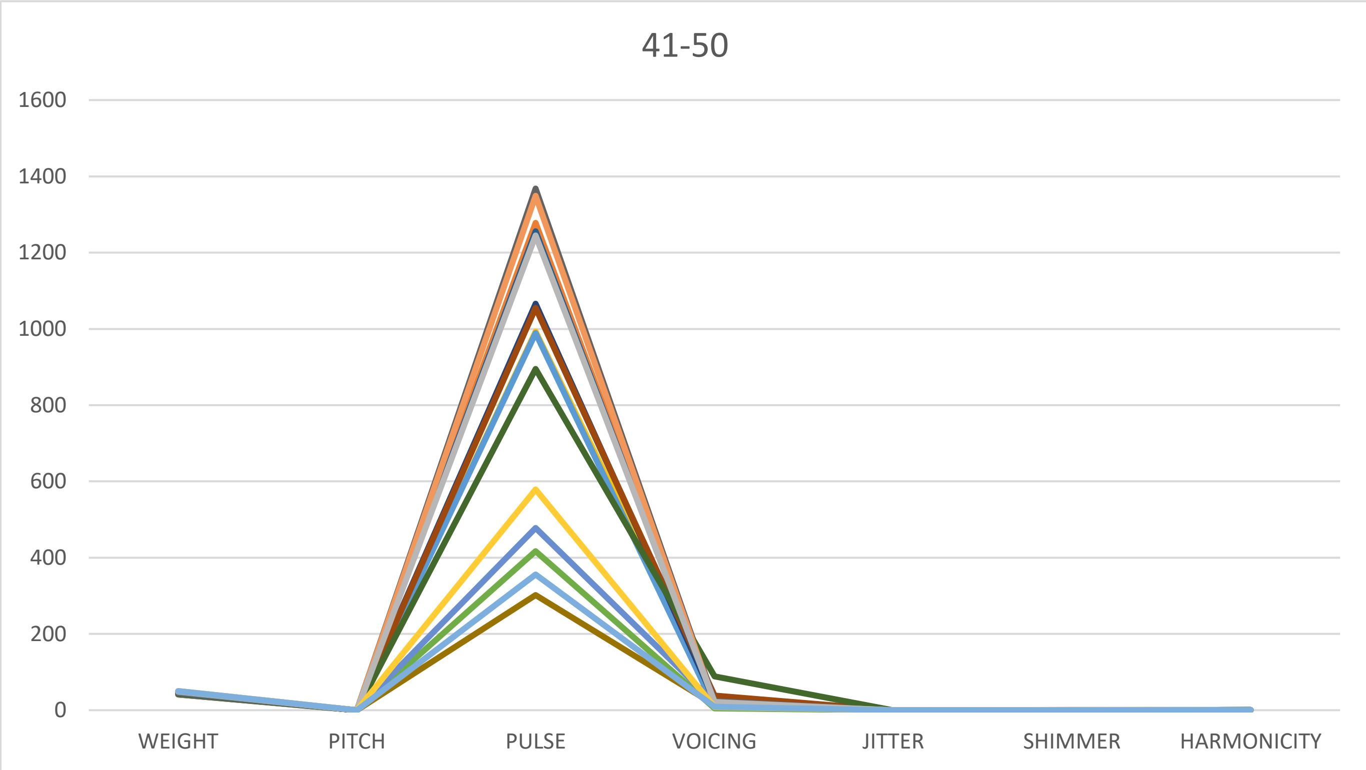
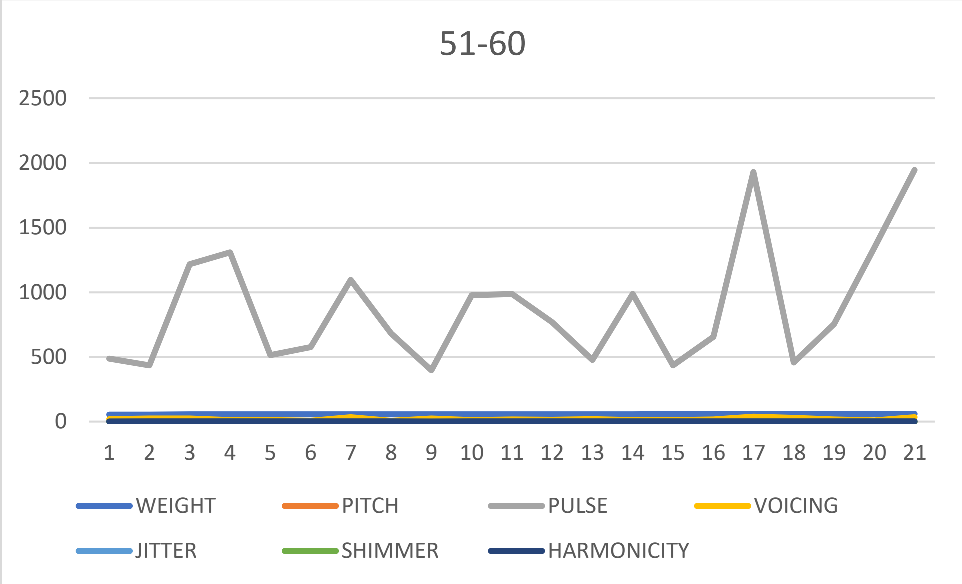


Table 8- Weight 51-60kgs, voice parameters readings.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
54	327.515 Hz	487	20	3.89%	14.82%	0.829829
54	327.515 Hz	436	23	3.86%	16.98%	0.774763
55	298.789HZ	1216	24	3.24%	17.89%	0.746085
55	327.515 Hz	1308	12	2.42%	13.35%	0.823768
55	327.515 Hz	514	12	3.21%	15.82%	0.756775
55	327.515 Hz	575	7	2.52%	16.63%	0.822467
56	239.465HZ	1095	35	3.24%	17.89%	0.746085
56	327.515 Hz	683	6	3.23%	14.49%	0.808762
56	327.515 Hz	397	25	3.98%	14.83%	0.723145
56	327.515 Hz	978	12	3.86%	16.98%	0.774763
56	327.515 Hz	987	17	3.35%	16.98%	0.786543
56	367.9807HZ	769	14	4.99%	17.87%	0.756447
57	327.515 Hz	476	18	3.76%	16.98%	0.775434
57	325.987HZ	987	11	4.99%	17.87%	0.756447
58	327.515 Hz	435	12	3.86%	16.98%	0.774763
58	376.986HZ	654	17	4.77%	17.34%	0.787456
59	310.289HZ	1930	38	3.97%	23.67%	0.736588
59	327.515 Hz	458	29	4.76%	16.98%	0.775644
59	367.988HZ	754	16	4.34%	16.97%	0.776542
60	310.289HZ	1345	10	2.44%	21.84%	0.746085
60	310.289HZ	1946	34	3.97%	21.84%	0.746085

Graph 11,12- Weight 61-70kgs, voice parameters variations.



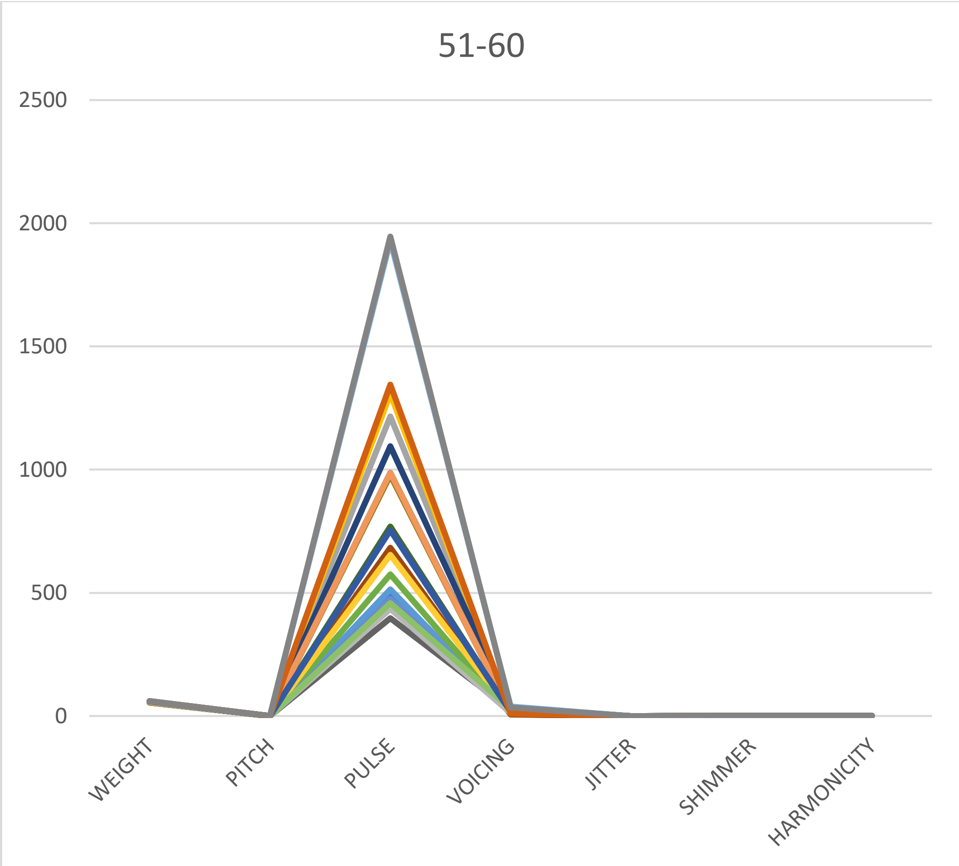
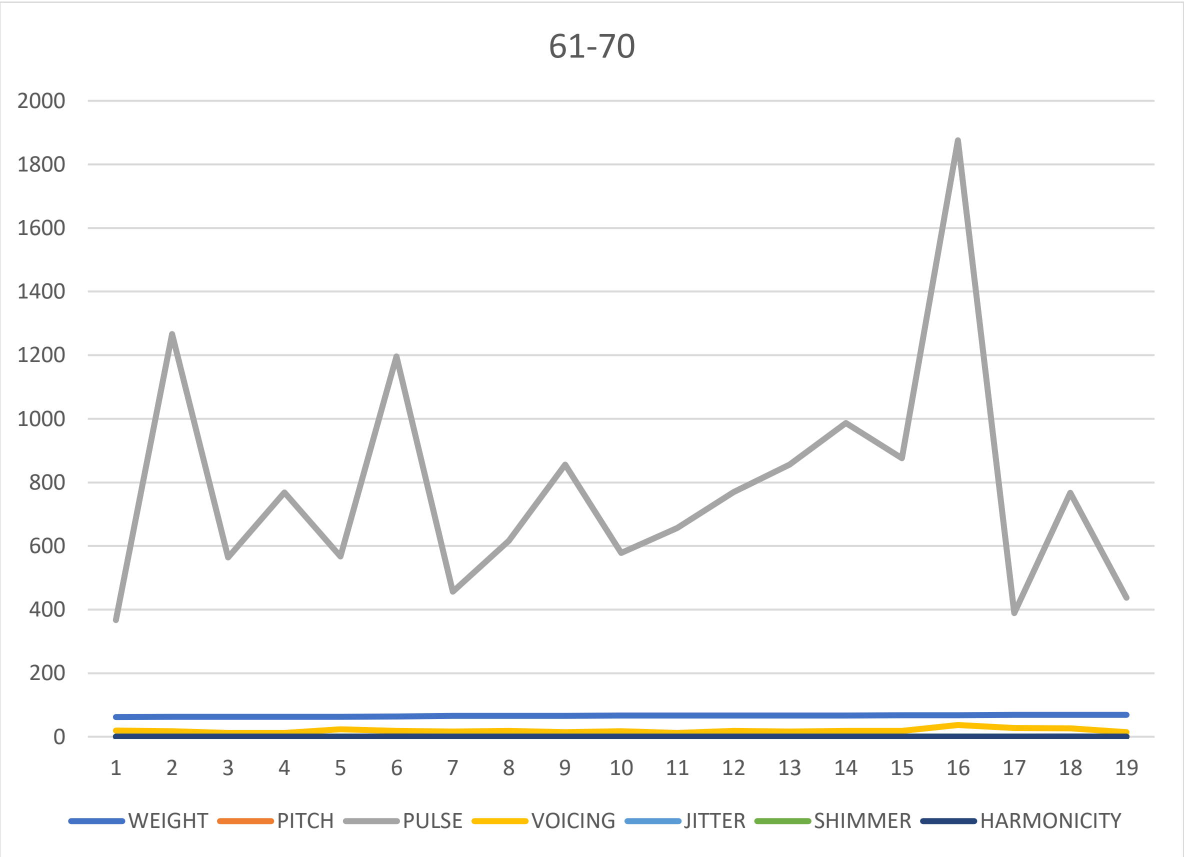


Table 9- Weight 61-70kgs, voice parameters readings.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
62	327.515 Hz	367	19	3.86%	14.78%	0.774763
63	345.895HZ	1267	17	3.97%	26.97%	0.765322
63	327.515 Hz	564	12	3.86%	16.98%	0.774763
63	327.515 Hz	768	12	4.78%	16.98%	0.776754
63	327.515 Hz	567	23	3.89%	16.98%	0.776543
64	327.515 Hz	1196	18	2.99%	14.41%	0.788512
66	327.515 Hz	456	16	3.67%	16.98%	0.776543
66	376.678HZ	616	18	4.87%	17.87%	0.779874
66	308.985HZ	856	14	3.98%	17.76%	0.775644
67	327.515 Hz	578	17	4.89%	16.98%	0.774763
67	345.876HZ	657	12	4.88%	17.87%	0.776547
67	376.877HZ	769	18	4.98%	16.97%	0.784535
67	365.097HZ	856	16	4.98%	17.98%	0.775428
67	309.865HZ	987	18	3.98%	16.76%	0.770546
68	327.515 Hz	876	18	4.67%	16.75%	0.776545
68	5436.987HZ	1876	37	4.85%	17.87%	0.776435
69	327.515 Hz	389	28	3.86%	15.78%	0.774763
69	234.987HZ	767	27	4.99%	17.87%	0.756447
69	376.986HZ	437	14	3.87%	16.67%	0.772347

Graph 10,11- Weight 61-70kgs, voice parameters variations.



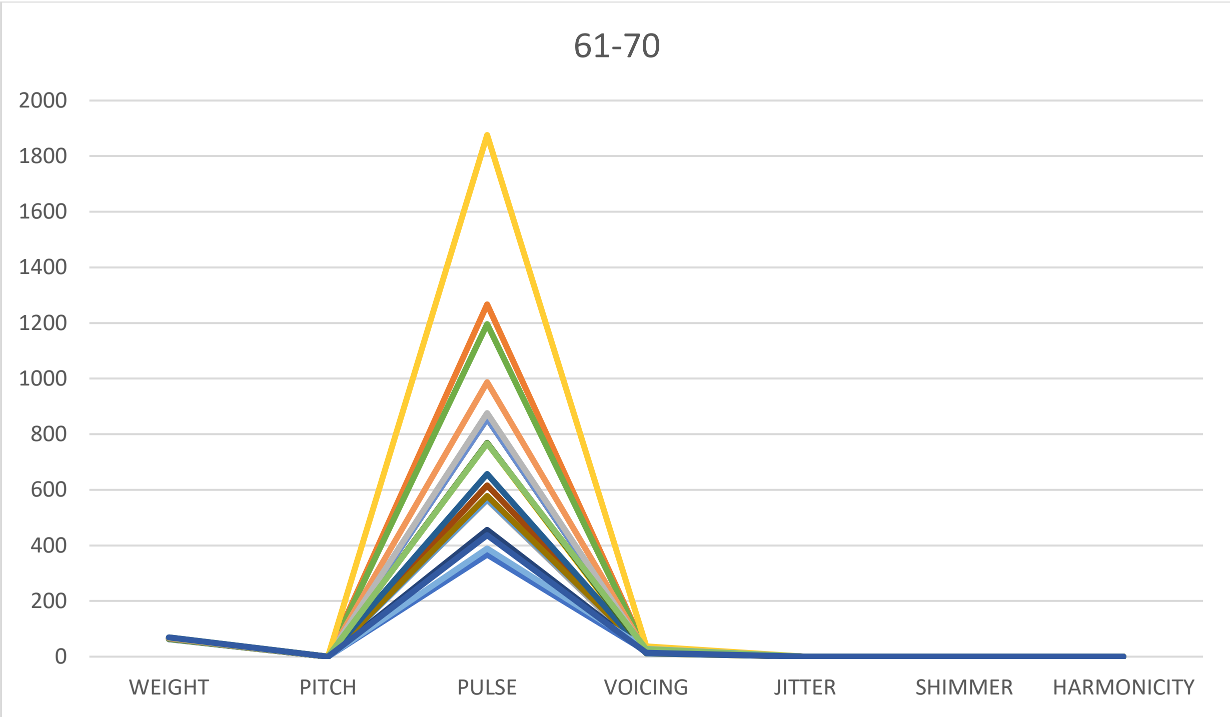
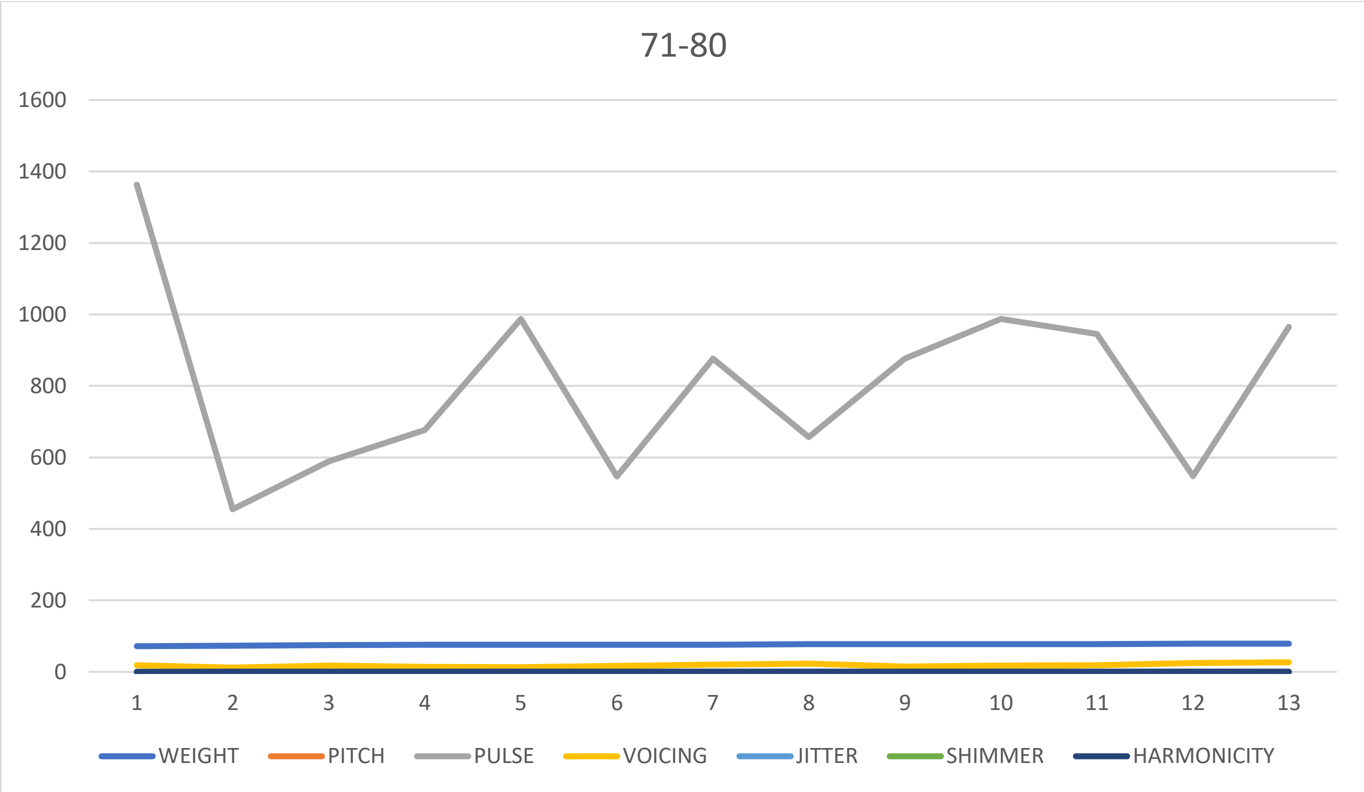


Table 10- Weight 71-80kgs, voice parameters readings

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
72	345.895HZ	1363	19	3.97%	14.58%	0.245839
73	327.515Hz	455	12	3.76%	17.47%	0.774354
75	327.515Hz	589	18	2.95%	14.78%	0.789688
76	327.515Hz	677	14	3.39%	16.98%	0.786954
76	357.954HZ	987	13	4.97%	17.87%	0.775643
76	301.976HZ	547	17	3.87%	17.87%	0.785322
76	398.097HZ	876	20	3.96%	17.87%	0.771232
78	327.515Hz	657	23	3.86%	16.98%	0.774763
78	387.987HZ	876	15	4.97%	17.87%	0.745333
78	365.857HZ	987	18	4.98%	16.45%	0.798765
78	387.765HZ	945	19	4.65%	16.45%	0.798453
79	456.987HZ	548	25	4.99%	17.87%	0.756447
79	265.878HZ	965	27	4.34%	16.45%	0.785456

Graph 12,13- Weight 71-80kgs, voice parameters variations.



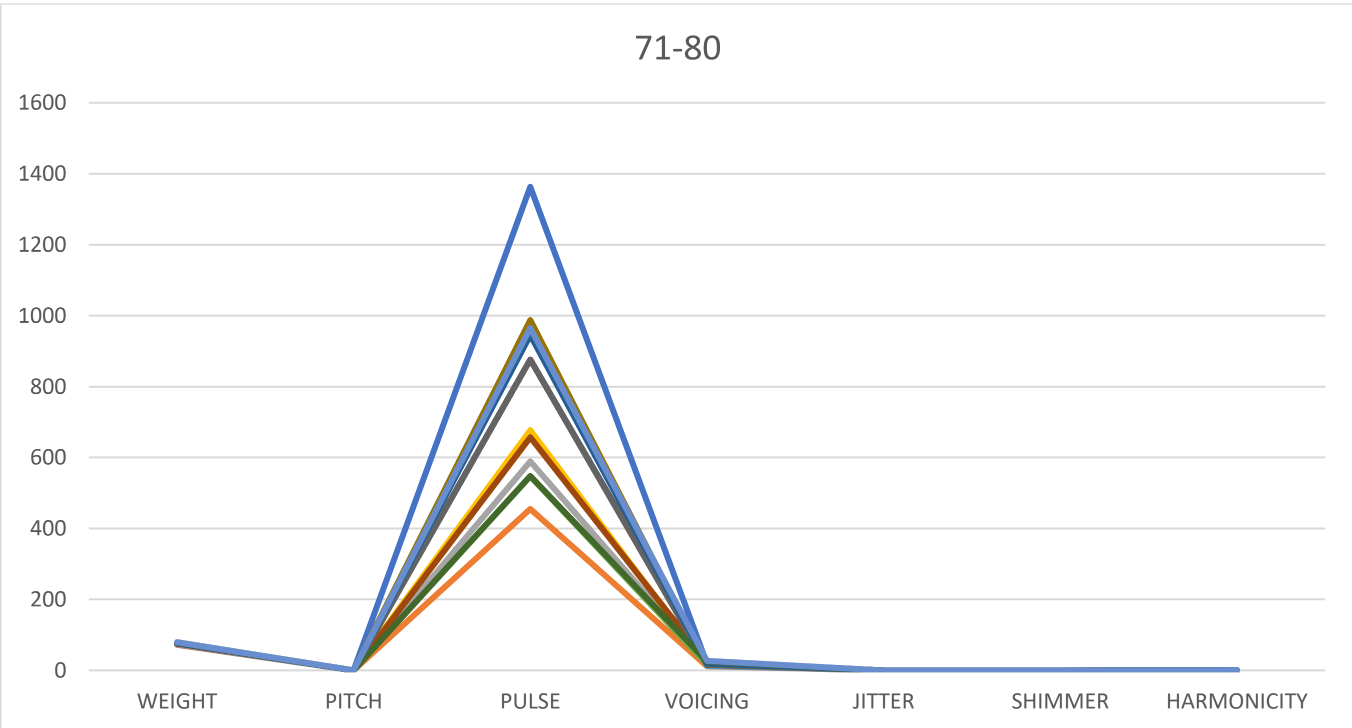
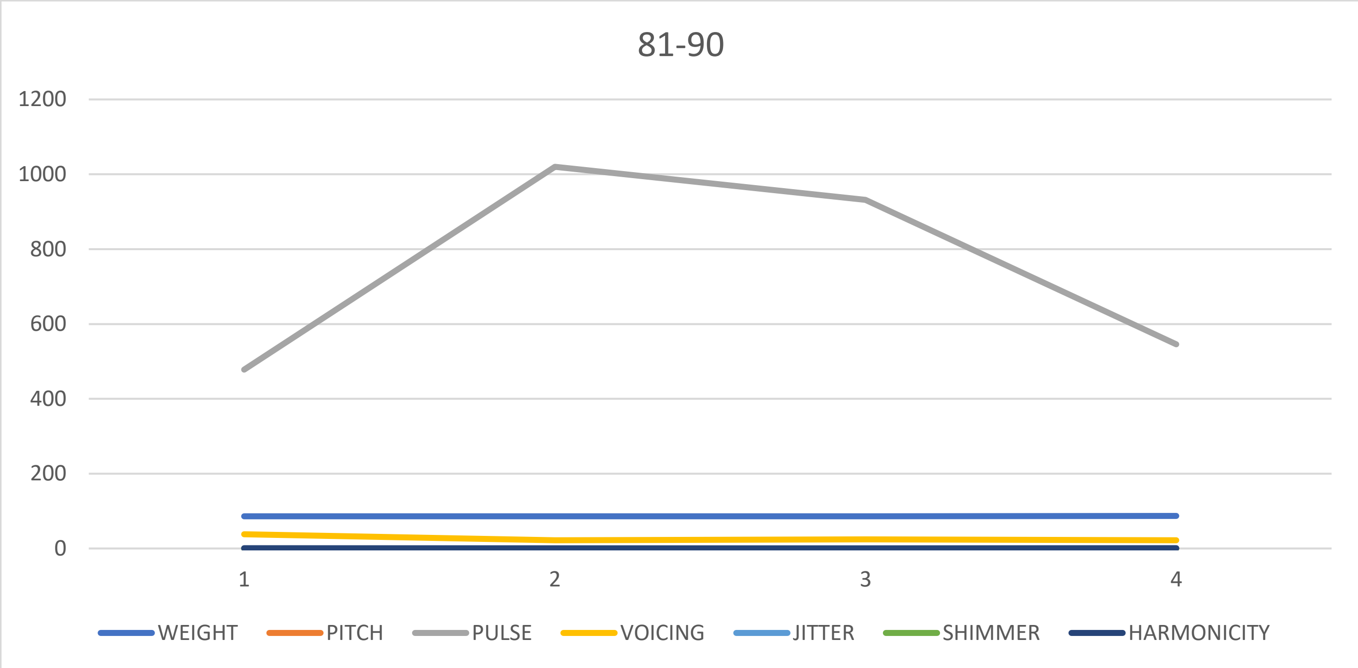


Table 11- Weight 81-90 kgs, voice parameters readings.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
86	327.515Hz	478	38	3.65%	16.56%	0.774763
86	327.515Hz	1020	22	3.76%	16.56%	0.775456
86	376.567HZ	932	24	4.54%	16.45%	0.785454
87	398.597HZ	546	22	4.43%	16.45%	0.798642
90	456.096HZ	1867	11	4.99%	17.87%	0.756447

Graph 14,15- Weight 81-90 kgs, voice parameters variations.



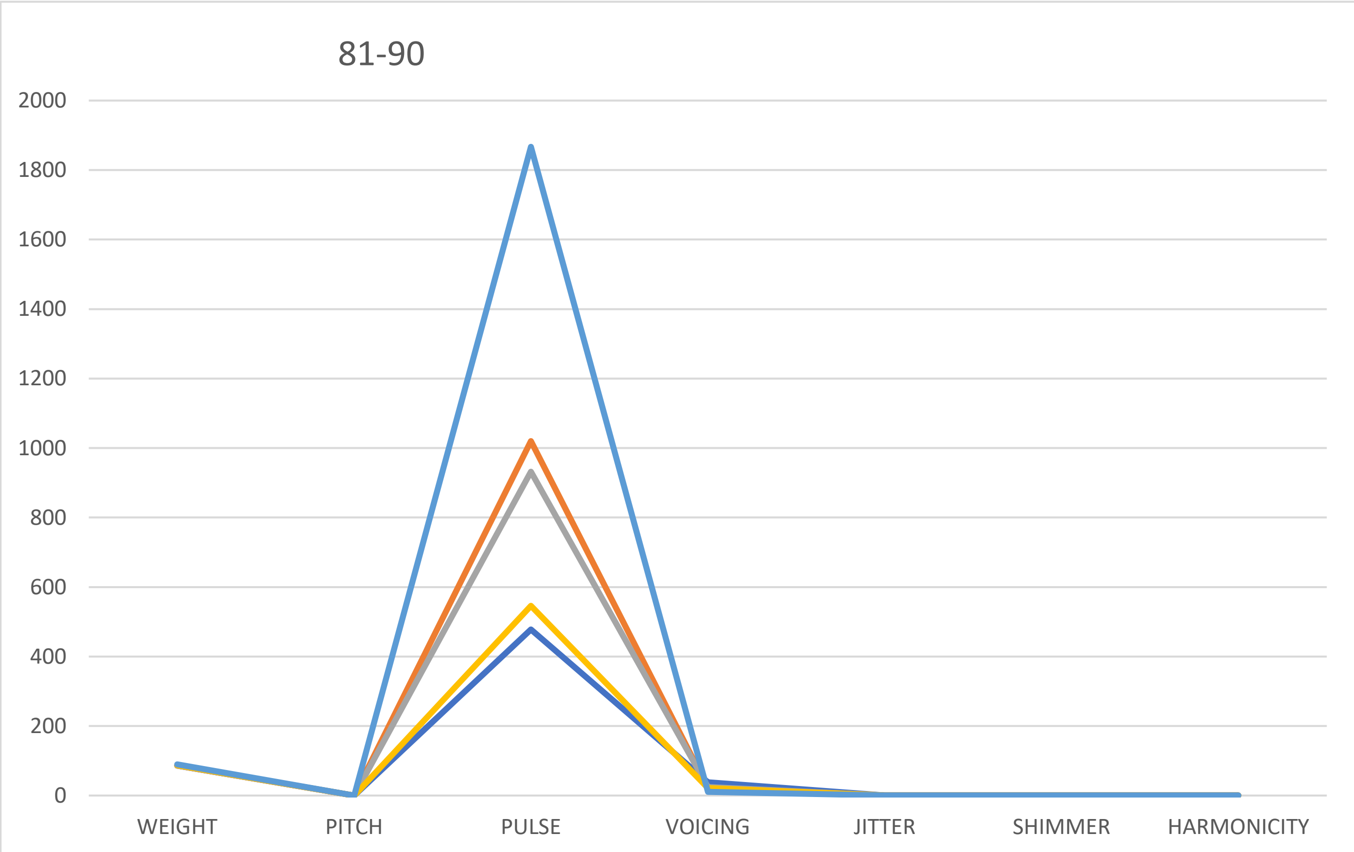
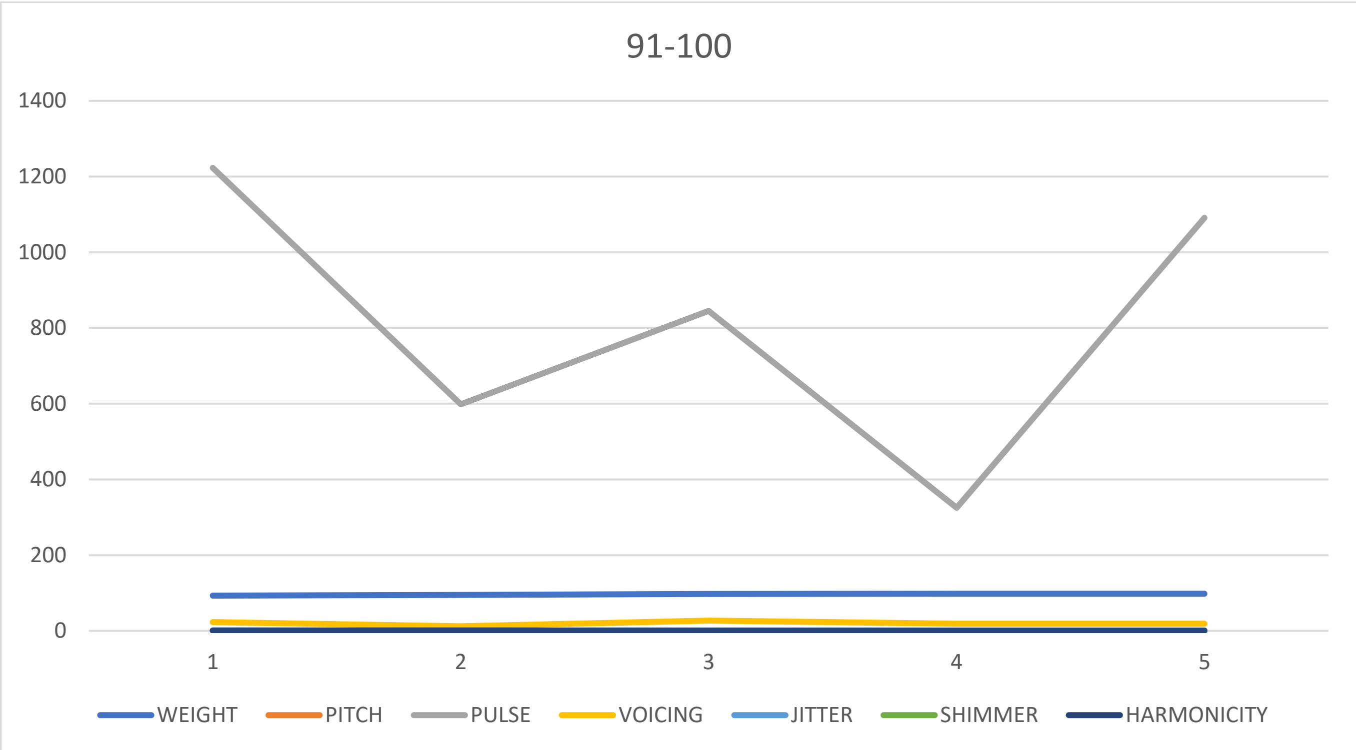


Table 12- Weight 91-100, voice parameters readings.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
93	327.515Hz	1223	23	3.76%	16.56%	0.775635
95	327.515Hz	598	12	4.87%	16.56%	0.774763
97	367.987HZ	845	27	4.76%	16.45%	0.879754
98	327.515Hz	325	19	4.87%	16.98%	0.756437
98	357.879HZ	1091	19	4.99%	17.87%	0.756447

Graph 16,17-- Weight 91-100, voice parameters variations.



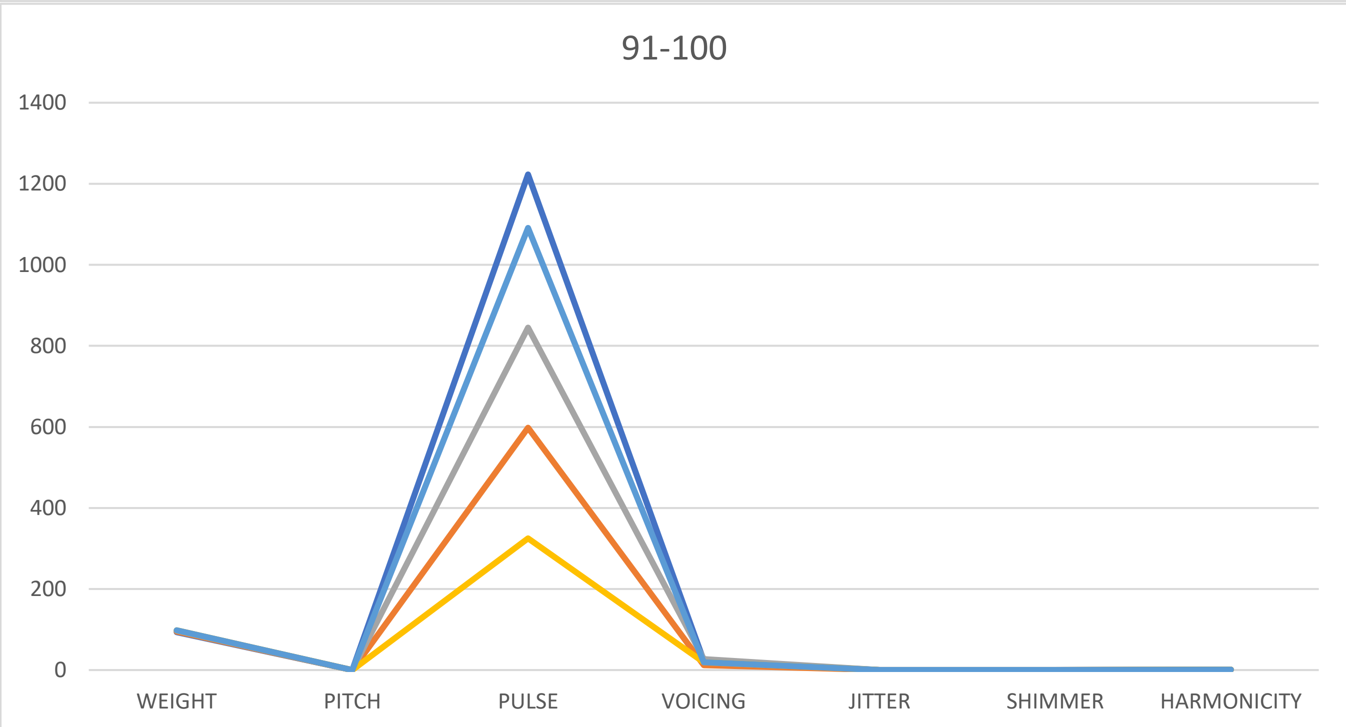
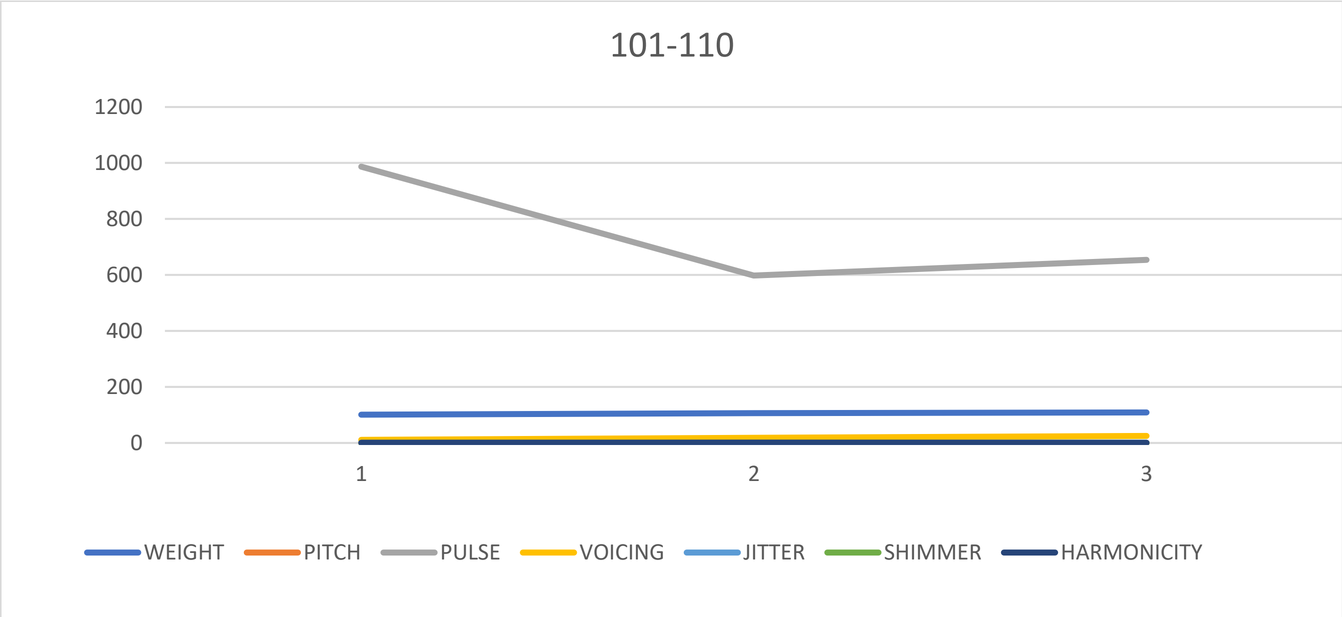
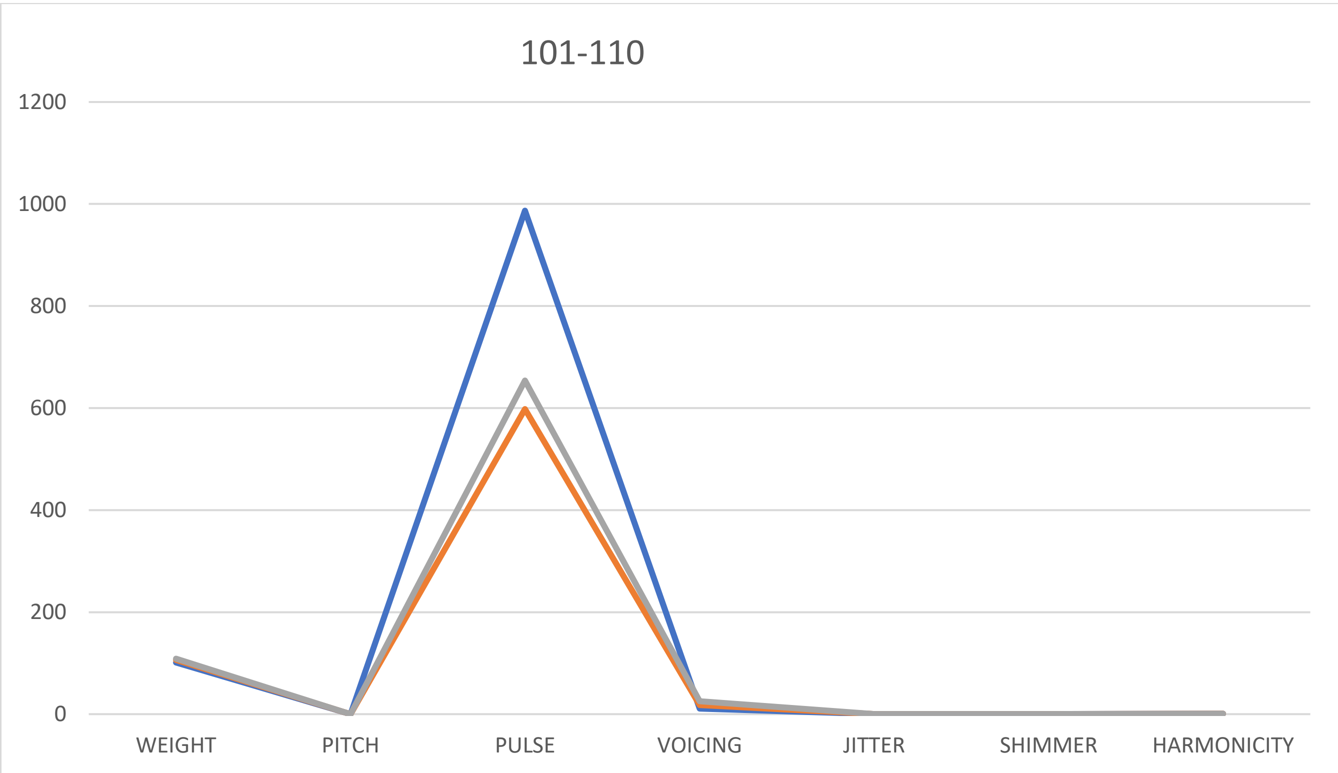


Table 13- Weight 100-110kgs, voice parameters readings.

WEIGHT	PITCH	PULSE	VOICING	JITTER	SHIMMER	HARMONICITY
101	327.515Hz	987	11	3.76%	16.56%	0.775643
106	327.515Hz	598	18	3.76%	16.56%	0.774763
109	345.098HZ	654	25	4.99%	17.87%	0.756447

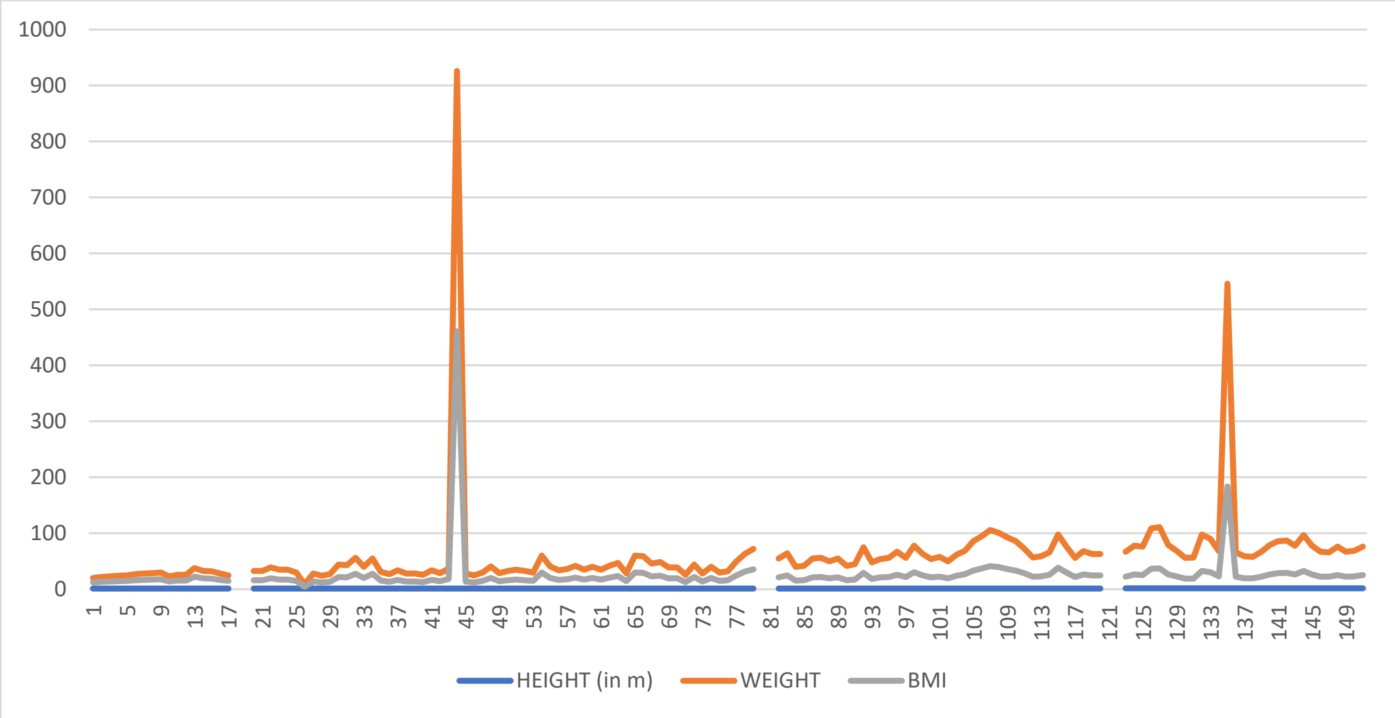
Graph 18-19-- Weight 100-110kgs, voice parameters variations.





From the above data analysis of different weight ranges that 20-30kgs, 31-40kgs, 41-50kgs, 51-60kgs, 61-70kgs, 71-80kgs, 81-90kgs, 91-100kgs, 101-110kgs, the voice parameters reading from the spectrographic analysis showed high variations in pulse, pitch and voicing, and moderate variations in the, jitter, shimmer, harmonicity.

Graph 20, BMI,Weight and height corelation



In the study, it was observed that there was a notable fluctuation in weight and BMI across five sample points, while height remained consistent.

5. Findings/ Discussion

This study aims to investigate the feasibility of estimating approximate height and weight through spectral analysis of voice. By analyzing voice samples collected from the study participants, seeking to ascertain whether it is possible to derive accurate estimations of height and weight based on vocal characteristics.

The findings of this study shed light on the complex relationship between anthropometric characteristics (height and weight) and voice parameters.

The spectrographic analysis of voice parameters across different height and weight ranges revealed significant variations in pulse, pitch, voicing, jitter, shimmer, and harmonicity.

Height Ranges Analysis:

Pulse and Pitch: Substantial variations in pulse and pitch were observed among individuals of different height ranges. This suggests that variations in vocal fundamental frequency and periodicity are influenced by anatomical factors such as vocal tract length. (Arsikere, 2013)

Voicing: Moderate variations in voicing were noted, indicating differences in the presence and duration of voiced segments within the voice signal among individuals of varying heights.

Jitter, Shimmer, and Harmonicity: These parameters exhibited moderate variations across height categories, suggesting some variability in vocal fold vibration irregularity, intensity variability, and harmonic structure. (5. Van Dommelen, (1995))

Weight Ranges Analysis:

Pulse, Pitch, and Voicing: High variations in pulse, pitch, and voicing were observed across different weight ranges, indicating substantial differences in vocal fundamental frequency, periodicity, and the presence of voiced segments among individuals with varying weights.

Jitter, Shimmer, and Harmonicity: Moderate variations were observed in jitter, shimmer, and harmonicity across weight categories, suggesting some variability in vocal fold vibration irregularity, intensity variability, and harmonic structure.

The observed variations in pulse, pitch, and voicing across height and weight ranges suggest that anatomical differences play a significant role in shaping vocal characteristics. Specifically, individuals with shorter or lighter stature may exhibit different vocal properties compared to those with taller or heavier builds, potentially due to differences in vocal tract length, vocal fold mass, or tension.

Moreover, the moderate variations in jitter, shimmer, and harmonicity across height and weight categories indicate additional factors influencing voice production. These parameters reflect aspects of vocal fold vibration irregularity, intensity variability, and harmonic structure, which may be influenced by physiological, biomechanical, or environmental factors beyond anthropometry.

6. Conclusion

In this study, the relationship between anthropometric characteristics, namely height and weight, and various voice parameters such as pulse, pitch, voicing, jitter, shimmer, and harmonicity was established. The analysis revealed significant variations in these parameters across different height and weight ranges, suggesting a complex interplay between anatomical factors and vocal characteristics.

The findings regarding height indicate that individuals with varying statures exhibit substantial differences in pulse and pitch, indicating that vocal fundamental frequency and periodicity are influenced by vocal tract length. Moreover, moderate variations in voicing suggest differences in the presence and duration of voiced segments among individuals of different heights. These results underscore the importance of considering anatomical variations in vocal production and its assessment. (Jaid, (2023))

Similarly, our analysis of weight ranges revealed high variations in pulse, pitch, and voicing, suggesting notable differences in vocal characteristics among individuals with different weights. This emphasizes the impact of factors such as vocal fold mass or tension on vocal fundamental frequency and the presence of voiced segments. Moderate variations in jitter, shimmer, and harmonicity across weight categories further highlight the complexity of voice production and the multifactorial nature of vocal variability.

Overall, the findings contribute to a deeper understanding of the influence of anthropometric characteristics on voice parameters. By elucidating the relationship between height, weight, and vocal characteristics, our study provides valuable insights for voice research and clinical practice. Understanding these relationships can aid in more accurate voice assessment, diagnosis, and treatment planning, particularly in fields such as speech pathology, forensic science, and biometrics. (Lass N. J., (1980))

Moving forward, further research is warranted to explore the underlying mechanisms driving the observed variations in voice parameters across different anthropometric categories. Investigating additional factors such as vocal fold morphology, respiratory function, and articulatory mechanisms may provide a more comprehensive understanding of vocal variability. Additionally, exploring the implications of these findings for voice-related disorders and interventions could contribute to the development of more personalized and effective treatment approaches.

In conclusion, our study underscores the importance of considering height and weight factors in voice assessment and highlights the need for continued research to advance our understanding of the complex relationship between anthropometric characteristics and vocal production.

7. Recommendations

Based on the findings of this study, several recommendations can be made for future research and clinical practice in the field of voice assessment and intervention:

1.More study: To better understand the underlying mechanisms causing the observed variances in voice characteristics across various anthropometric categories, more study is required. A more thorough understanding of vocal variability might be obtained by looking into other variables such articulatory mechanics, respiratory function, and vocal fold morphology. (Künzel, 1989)

2.Clinical Application: Voice assessment procedures in clinical practice can be improved by the knowledge gathered from this study. To increase the precision of voice diagnostic and treatment planning, medical professionals—especially speech pathologists and otolaryngologists—should think about integrating anthropometric measures like weight and height into their evaluations.

3. Treatment Personalisation: Developing more individualised and successful treatment plans for voice-related diseases can be facilitated by knowledge of the connection between anthropometric traits and vocal parameters.

8. REFERENCES

1. https://corpus.eduhk.hk/english_pronunciation/index.php/1-2-introduction-of-praat/
2. (Bruckert, (2006)). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 83-89.
3. Gonzalez, J. (2003). Estimation of speakers' weight and height from speech: A re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and motor skills*, *96*(1), 297-304.
4. Lass, N. J., & Colt, E. G. (1980). A comparative study of the effect of visual and auditory cues on speaker height and weight identification. *Journal of Phonetics*, *8*(3), 277-285.
5. (5. Van Dommelen, (1995)). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and speech*, *38*(3), 267-287.
6. Dusan, S. (2005). Estimation of speaker's height and vocal tract length from speech signal. In *Interspeech* (pp. 1989-1992).
7. Lass, N. J., Hendricks, C. A., & Iturriaga, M. A. (1980). The consistency of listener judgments in speaker height and weight identification. *Journal of Phonetics*, *8*(4), 439-448.
8. Lass, N. J., Phillips, J. K., & Bruchey, C. A. (1980). The effect of filtered speech on speaker height and weight identification. *Journal of Phonetics*, *8*(1), 91-100
9. Lass, N. J., Dicola, G. A., Beverly, A. S., Barbera, C., Henry, K. G., & Badali, M. K. (1979). The effect of phonetic complexity on speaker height and weight identification. *Language and Speech*, *22*(4), 297-309.
10. Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight?. *Phonetica*, *46*(1-3), 117-125
11. Jaid, U. H., & Abdulhassan, A. K. (2023). Fuzzy-Based Ensemble Feature Selection for Automated Estimation of Speaker Height and Age Using Vocal Characteristics. *IEEE Access*.
12. Poorjam, A. H., Bahari, M. H., & Van Hamme, H. (2015, March). Speaker weight estimation from speech signals using a fusion of the i-vector and NFA frameworks. In *2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP)* (pp. 118-123). IEEE.
13. (Arsikere H. L., 2013). Automatic estimation of the first three subglottal resonances from adults' speech signals with application to speaker height estimation. *Speech Communication*, *55*(1), 51-70.
14. (González, 2004) Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of phonetics*, *32*(2), 277-287.

9-Bibliography

- Van Dommelen, W. A. ((1995)).
- Arsikere, H. L. (2013).
- Bruckert, L. L. ((2006)).
- Dusan, S. ((2005)).
- Gonzalez, J. ((2003)).
- González, J. (. (2004).
- Jaid, U. H. ((2023)).
- Künzel, H. J. (1989).
- Lass, N. J. ((1980)).
- Lass, N. J. (1979).
- Poorjam, A. H. ((2015, March)).