A Study of the Variation of Acoustic Properties of Signals Received by Used Mobile Phones

Somarathne K.A.S.K.[1], Kalansooriya C.M.[2] and Gamage N.G.S.S. [1*]

[1]Department of Physics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka

[2]Electro Technology Laboratory, Industrial Technology Institute, 363, Bauddhaloka Mawatha, Colombo 07, Sri Lanka

shanthagamage@sci.sjp.ac.lk

1. ABSTRACT

This study investigated the variation of acoustic properties of signals received by mobile devices, focusing on the quality of microphones and speakers. The study was motivated by the economic crisis in Sri Lanka, which has made it difficult for people to afford new mobile phones. Old mobile phones undergo wear and tear on the microphones and speakers with time. Voice recordings were collected with low background noise using a sound level meter and an internal recording app. One mobile phone acted as the receiver, recording voice signals transmitted by other mobile phones acting as callers. These received voice signals were collected under normal environmental conditions. Results indicated that the mean Equivalent Continuous Sound Level (LAeq) for Brand A and Brand B phones was nearly equivalent to the LAeq of the original voice. The minimum and maximum LAeq values for the recorded voices of Brand A (-10.34 and 62.26, respectively) and Brand B 01 (-13.38, 64.32, respectively) covered a comparable range as the original voice (-14.84 and 60.87, respectively). The Wilcoxon signed-rank test affirmed that there was no significant difference between the original voice and the recorded voice of Brands A, B, and C, suggesting that mobile phones possess a good capability to capture sound pressure levels. Additionally, there were no significant differences between the recorded voices of Brand A and Brand B and the received voices from others. The study suggests that while Brands A and B exhibit strong performance in capturing and transmitting voice with minimal differences, Brands C, D and E experience signal loss or degradation, along with less durability of the microphone and speaker components. Factors like formant, fundamental frequency, and shimmer underscore the intricacies of voice recording and transmission. In conclusion, this research asserts that the microphones of Brands A, B, and C mobile phones demonstrate good durability in recognizing human voices, and Apple and Sony mobile phones can be effectively used for voice recognition in call reception.

Keywords - Equivalent Continuous Sound Level, Acoustic property, Microphone, Voice parameters

2. INTRODUCTION

People use various smartphones which are available in the market with different names. The performance of these mobile phones is changed according to the prices, brand names, and features. Mobile phones are mainly used for communication with other people. Microphones and speakers are the main tools that are used for communication. Therefore, the performance of these devices is more important. Currently, Sri Lanka's economic crisis renders the purchase of new mobile phones unaffordable for many individuals. Due to this reason, people try to buy used phones. Mobile phones were utilized for a long period of time, there are certain issues caused by the microphones and speakers of mobile phones. The speakers and microphones in mobile phones, like many electronic components, can experience wear and tear over time. It is obvious that when using a telephone, the receiver does not hear the actual sound the caller is making. Some changes can be seen in the receiver's voice. Therefore, the main objective of this research project was to investigate the reason for this difference in voice between the caller and the receiver by doing speaker quality analysis and microphone quality analysis. To analyze the differences between the voice characteristics of the recorded voice from a mobile phone and the received voice signal from the same mobile phone with the original voice, mainly the sound pressure level of the voice is utilized.

An acoustic system can be modeled both in the time domain and the frequency domain. Acoustic parameters in the frequency domain can be derived from the domain analysis of sound, various parameters are derived from the spectral representation of a signal. These parameters provide information about the distribution of energy across different frequency components [1].

Time domain parameters such as the jitter, the shimmer, signal-to-noise ratio (SNR), and the harmonic-to-noise ratio (HNR) were used and for frequency domain analysis fundamental frequency was used [2] – [5]. Jitter is a measure of the cycle-to-cycle variations in the fundamental frequency of a periodic signal, such as the human voice. For voice detection, formants, shimmer, jitter, HNR, and CCP are measured to compare the recorded voice and received voice from the mobile phone [6].

Sound pressure level (SPL) is a measure of the sound pressure level relative to a reference level. It quantifies the intensity or loudness of a sound. SPL is typically expressed in decibels (dB) and is often used to measure and compare the acoustic power or intensity of different sounds. It represents a logarithmic measure of the ratio between the sound pressure of a particular sound wave and a reference sound pressure level. The reference sound pressure level commonly used is 20 micro-Pascals (µPa), which is approximately the threshold of human hearing at a frequency of 1 kHz [7].

SPL correlates with the perceived loudness of a sound. Generally, as SPL increases, the sound is perceived as louder, and as SPL decreases, the sound is perceived as softer. This can be measured using a sound level meter, which consists of a microphone for capturing the sound and a meter that displays the SPL reading. These measurements are often weighted using frequency weighting filters (such as A, B, or C weighting) to account for the varying sensitivity of the human ear to different frequencies.

When measuring the equivalent sound pressure level (SPL) of the recorded voice sample, the presence of background noise can impact the overall measurement. Noise can affect the voice or sound signal in recording stage, transmission stage, and receiving stage [8].

Different frequency weightings are commonly applied to measurements in acoustic analysis to approximate the human ear's frequency sensitivity. Some commonly used frequency weighting curves include A-weighting, C-weighting, and Z-weighting. A-weighting is a frequency weighting curve designed to approximate the frequency response of the human ear at moderate sound levels. It emphasizes frequencies in the range of 500 Hz to 10 kHz while attenuating low and high frequencies. It is widely used in noise measurements and standards to assess noise levels experienced by humans. A-weighting is most sensitive to the 1-5KHz range [9].

Approximately 30 years ago, the Union of European Phoneticians (UEP) made efforts to establish a standardized method for measuring voice range profiles. They recommended that these measurements be conducted at a microphone distance of 30 cm from the mouth using a sound level meter with the standard A-weighting frequency protocol (Schutte and Seidner, 1983) [10]. Since then, the 30 cm distance has been widely accepted as the standard for voice measurements. However, it should be noted that in some studies, alternative distances have been used, leading to difficulties when comparing results across different studies.

3. METHODOLOGY

Speech data was collected in two different environmental conditions at the Industrial Technology Institute (ITI), Colombo. A voice signal was played using ODEON to collect the voice data. The output voice signal from ODEON was amplified by Power Amplifier type 2734 (Bruel & Kjaer). Every voice sample was captured in a noise-controlled environment, ensuring the ambient noise level remained below 40 dB [9]. The voice signal received by the speaker was recorded from an internal sound recorder application by placing the smartphone approximately 30 cm away from the omnidirectional speaker, positioned directly in front of it. This protocol was carried out for 10 mobile phones to record the voice signal (Brand A, B, C, D and E). At the same time, the sound level meter (Bruel & Kjaer type 2250) was placed to detect the voice signal. A weighted filter was utilized to record the LAeq of the voice signal.

One of the mobile phones worked as a receiver-mobile phone kept in an outside environment of ITI (55dB) while other devices worked as caller-mobile phones. The voice signal from the omnidirectional speaker was transmitted through mobile communication to record the voice date from the sound level meter and call recorder app.

The voice note recordings were initially in OGG and MP4 formats performing with a sampling rate of 44.1 kHz and 32 bits per[11], utilizing a mono channel. Before extracting the features, the recordings underwent normalization. The feature extraction of recorded voice was analyzed by the Praat software's latest version. All the data were analyzed using SPSS 26 (IBM SPSS Statistics, Armonk, NY). The Wilcoxon rank sum tests were conducted to analyze the following sound level meter measurements.

4. RESULTS AND DISCUSSION

The mean LAeq values of Apple and Sony are 30.4958 dB and 30.017 dB, which is close to the mean LAeq value of the original voice at 30.2194 dB. The minimum and maximum LAeq values of the recorded voice of Brand A (-10.34 and 62.26, respectively) and Brand B-01 (-13.38, 64.32 respectively) cover a comparable range as the original voice (-14.84 and 60.87, respectively), indicating the microphone's capability to capture a similar range of sound levels.

The Wilcoxon rank sum test is a nonparametric statistical test used to assess whether there is a significant difference between the sound pressure of recorded voices from different mobile phones and the original voice. Wilcoxon signed-rank test confirmed, there was no significant difference between the original voice and recorded voice of Brand A (p=0.668), Brand B 01 (0.454), Brand B 02 (0.044), Brand C 01 (0.264), and Brand C 02 (0.532), which means these mobile phones have good capability to capture the sound pressure level.

Brand A recorded voice was transmitted to the receiving mobile phones such as Brands B 01, C 01, and E, the respective p-values of these original voice and received voice combinations are 0.879, 0.823, and 0.085. There is no significant difference between the Brand A recorded voice and the receiver voice. Also the Brand B 01, there is no significant difference between the recorded voice and the received voice. Brand B 01 recorded voice was transmitted to the receiving mobile phones such as Brand A (p = 0.311), E (0.40), C 01 (.249), and D 03 (.288). Brand B 02 recorded voice was transmitted to the receiving mobile phones such as Brand D 03 (0.131), C 02 (0.022), D 02 (0.427), and D 01 (0). The received voices of Brand D 02 and C 02 are significantly different from the recorded voices of Brand B 02 due to Signal Loss or Degradation that may occur due to signal strength issues. By considering the Brand C 01 (Brands A (0.922), E(0.598), B 01(0.416)) and C 02 (Brand D 03 (0.043), D 04 (0.000), D 01 (.070)recorded voice, there is no significant difference between the recorded voice of Brand C 01 and received voice of other mobile phones. Brand D mobile phones show, that there is a significant difference between the original voice and the received voice Tables 01 and 02 show the variation of p and z values of LAeq of recorded voice of mobile phones, original voice and received voice.

Table 01 - The Table of p and z values of LAeq variation of recorded voice of mobile phones and original voice

recorded Voice	**Z**	**p**
Brand A 01- O. voice (original voice)	-.3725b	0.696
Brand A 02- O. voice	-.429c	0.668
Brand B 01 - O. voice	-2.010c	0.044
Brand B 02 - O. voice	-.748c	0.454

Brand C 01 - O. voice	-1.117b	0.264
Brand C 02- O. voice	-.625b	0.532
Brand D 01 - O. voice	-2.582b	0.01
Brand D 02 - O. voice	-1.814c	0.07
Brand D 03 - O. voice	-2.117b	0.034
Brand E - O. voice	-2.117b	0.034

Table 02 - The Table of p and z values of LAeq variation of recorded voice of mobile phones and received voice of mobile phones

Receiver	Caller mobile phone									
	Brand A 02	Brand A 01	Brand B 02	Brand B 01	Brand C 01	Brand C 02	Brand D 02	Brand D 01	Brand D 03	Brand E
Brand A 02		0.67	0.231	0.623	0.085	0.089	0.392	0.12	0.24	0.879
Brand A 01	0.72		0.41	0.51	0.21	0.14	0.324	0.35	0.26	0.087
Brand B 02	0.23	0.13		0.12	0.1	0.022	0.427	0.027	0.05	0.08
Brand B 01	0.31	0.41	0.32		0.053	0.071	0.27	0.301	0.201	0.41

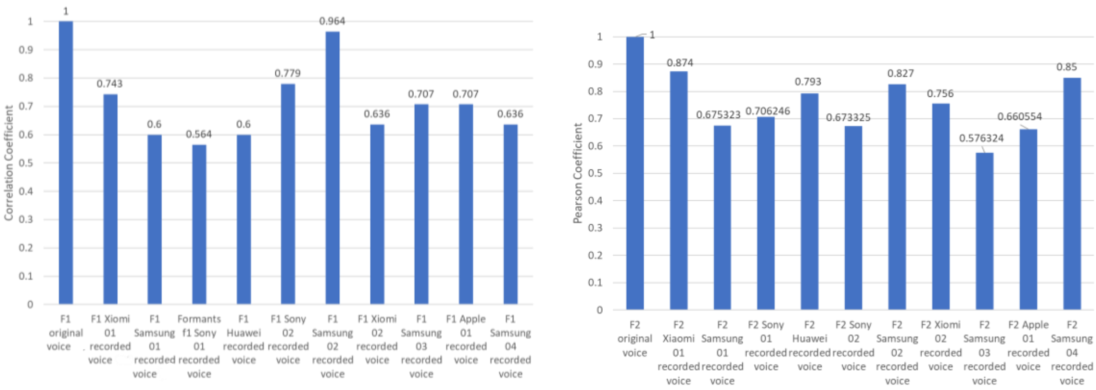

Figure 01: Correlation coefficient of recorded voices for formants. (a) Correlation Coefficients for formant 01, (b) Correlation Coefficients for formant 02

According to Figure 01, all the correlation coefficients are in the strong or mild range. Therefore, changes of the F1 and F2 are small for the recorded voice for the mobile phones compared to the original voice. For Brands A, B 01, and C 01 recorded voice from the microphone has a significant correlation with the received voice from different mobile phones. For other mobile phones, it is difficult to find the direct connection between the recorded voice and received voices because some received voices and recorded voices have a significant correlation, but some are not significant.

The fundamental frequency of the original voice is 120 Hz. The recorded voice of all mobile phones shows a fundamental frequency of around 120 Hz. But it is slightly greater than the original voice. Comparing The recorded voice of the phone and the received voice F0 is considerably high value. It varies from 120 to 200 Hz. These results indicate that when audio files are compressed for storage or transmission (as in the case of phone calls or voice messages), some frequency information may be lost. This loss of data can affect the perceived fundamental frequency of the recorded voice and Different speakers including those in mobile phones may have different frequency responses. Considering the HNR ratio of the original voice is 12. The HNR of recorded voices is around 7 which means background noise is involved for mobile phones when recording voice. The shimmer of the original voice is 10%. but the received voice of other mobiles has a high shimmer compared to the original voice, therefore the amplitudes of the received voice are high values.

5. CONCLUSION

The study indicates that while some mobile phones demonstrate good performance in capturing and transmitting voice with minimal differences, others may experience signal loss or degradation. Factors such as formant correlation, fundamental frequency variations, and shimmer differences highlight the complexity of voice recording and transmission. The microphones of Brands A, B and C mobile phones have good durability to recognize the human voice and Brand A and B mobile phones can be used for voice recognition for receiving calls.

6. REFERENCE

[1] J.-S. Liénard and M.-G. Di Benedetto, "Effect of vocal effort on spectral properties of vowels," *J Acoust Soc Am*, vol. 106, no. 1, pp. 411–422, 1999.

[2] F. Asci *et al.*, "Machine-learning analysis of voice samples recorded through smartphones: The combined effect of ageing and gender," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–17, 2020, doi: 10.3390/s20185022.

[3] C. T. Ferrand, "Harmonics-to-noise ratio: An index of vocal aging," *Journal of Voice*, vol. 16, no. 4, pp. 480–487, 2002, doi: 10.1016/S0892-1997(02)00123-6.

[4] I. R. Titze *et al.*, "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization," *J Acoust Soc Am*, vol. 137, no. 5, pp. 3005–3007, 2015.

[5] R. P. Baughman and R. G. Loudon, "Sound spectral analysis of voice-transmitted sound," *American Review of Respiratory Disease*, vol. 134, no. 1, pp. 167–169, 1986.

[6] F. Asci *et al.*, "Machine-learning analysis of voice samples recorded through smartphones: The combined effect of ageing and gender," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–17, Sep. 2020, doi: 10.3390/s20185022.

[7] G. Fava, G. Oliveira, M. Baglione, M. Pimpinella, and J. B. Spitzer, "The use of sound level meter apps in the clinical setting," *Am J Speech Lang Pathol*, vol. 25, no. 1, pp. 14–28, Feb. 2016, doi: 10.1044/2015_AJSLP-13-0137.

[8] L. R. Rabiner, "Applications of Voice Processing to Telecommunications."

[9] H. Šrámková, S. Granqvist, C. T. Herbst, and J. G. Švec, "The softest sound levels of the human voice in normal subjects," *J Acoust Soc Am*, vol. 137, no. 1, pp. 407–418, 2015, doi: 10.1121/1.4904538.

[10] H. K. Schutte, "UEP-recommendation for the standardisation of the voice area measurements/phonetography[EMPFEHLUNG DER UEP: STANDARDISIERUNG STIMMFELDMESSUNG/PHONETOGRAPHIE] Singing Voice physiology View project voice disorders View project," 1982. [Online]. Available: https://www.researchgate.net/publication/282203314

[11] J. Delgado-Hernández, N. M. León-Gómez, L. M. Izquierdo-Arteaga, and Y. Llanos-Fumero, "Cepstral analysis of normal and pathological voice in Spanish adults. Smoothed cepstral peak prominence in sustained vowels versus connected speech," *Acta Otorrinolaringologica (English Edition)*, vol. 69, no. 3, pp. 134–140, 2018.