2017

Comparison of various technologies in the acquisition of speech samples for transcription purposes: listener perception vs. acoustic analysis of signal quality

Alexa Klimes
University of Northern Iowa

COMPARISON OF VARIOUS TECHNOLOGIES IN THE ACQUISITION OF SPEECH SAMPLES FOR TRANSCRIPTION PURPOSES: LISTENER PERCEPTION VS. ACOUSTIC ANALYSIS OF SIGNAL QUALITY

A Thesis Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors

Alexa Klimes

University of Northern Iowa

May 2017

This Study by: Alexa Klimes

Entitled: Comparison of Various Technologies in the Acquisition of Speech Samples for Transcription Purposes: Listener Perception vs. Acoustic Analysis of Signal Quality

has been approved as meeting the thesis or project requirement for the Designation University Honors

________	________________________________
Date	Dr. Lauren Nelson, Ph.D., CCC-SLP, Honors Thesis Advisor, Communication Sciences and Disorders
________	________________________________
Date	Dr. Jessica Moon, Director, University Honors Program

Abstract

The purpose of this study was to investigate the quality of recordings obtained with a dedicated recording device in comparison to more readily available devices. Professionals in the field of communication disorders need to identify and transcribe specific details about individuals' speech and need high quality recordings for this purpose (Louko & Edwards, 2001). The dedicated recording devices that provide high quality signals are not readily available and can be expensive. The available research on this topic focuses on analysis of voice characteristics such as fundamental frequency, amplitude, and stability of amplitude and frequency (Lin, Hornibrook, & Ormond, 2012; Vogel & Maruff, 2008). This study extended this research to acquisition of samples for phonetic transcription of speech sounds. This research addressed two questions: (1) What combination of microphones and recording devices provides the clearest speech sample in terms of acoustic analysis? (2) Do listener preferences align with the quality of the recordings based on the acoustic analysis? Participants in this study included 5 adults who provided speech samples for the study and 20 adults who served as listeners and judged the quality of the recorded samples. The participants included both males and females, ages 19-23. Participants had hearing within normal limits, were native speakers of English, and were free of any speech or language disorders at the time of the study. The acoustic analysis for this study yielded information about the signal-to-noise ratio (SNR) for each device and the amount of peak clipping across the samples. This research provides information about the quality of recorded speech obtained from a dedicated digital audio recorder, a laptop computer, and an iPad. The findings showed that peak clipping was not a factor in selecting a device because this occurred only one time across all of the samples. The results for the SNR showed the iPad technology combination had the highest SNRs but also the largest measurement variability. The Marantz

technology combination had the lowest SNRs with the least amount of measurement variability. A different pattern emerged from the listener perception data. The listeners rated samples from all three of the devices approximately equally with regard to presence of noise in the signal and the signal clarity. Taken collectively, the results of this study suggest readily available technology, such as tablets and laptop computers, can be used to obtain high quality recordings of speech.

Introduction

Speech-language pathologists (SLPs) “work to prevent, assess, diagnose, and treat speech, language, social communication, cognitive-communication, and swallowing disorders in children and adults” (*Speech-Language Pathologists*, n.d.). The present study focused primarily on one aspect of speech-language pathology; the assessment of speech sound disorders through the acquisition of speech samples. Speech-language pathologists analyze speech samples in order to understand the specific difficulties and errors produced in speech. To do this, speech samples are carefully transcribed to note each and every detail of the sample, including the actual sounds the speaker says (Small, 2016). In order for the transcription to be as accurate as possible, it is essential that the speech sample is acquired in a way that presents the SLP with a high quality recording. While transcriptions can be made live, having a recording is more desirable. The ability to listen to a recording multiple times allows for greater reliability and accuracy of the transcription and also removes the stress of keeping pace with the speaker’s rate as is necessary when done in a live setting (Louko & Edwards, 2001).

Historically, speech samples have been acquired using dedicated digital audio recorders. With an increased use of modern technology in people’s lives, some research has been done in utilizing accessible technology, such as laptop computers, tablets and cellular devices to analyze clients’ speech and language (Lin et al., 2012; Vogel & Maruff, 2008; Vogel, Rosen, Morgan, & Reilly, 2014). The available research, however, focuses on voice analysis as opposed to acquiring samples for transcription of speech sounds (e.g., phonetic transcriptions). Transcribing speech is a perceptual process (Louko & Edwards, 2001) and therefore, more research is needed in evaluating the perceived quality of such recordings to determine whether or not they are adequate for use in the transcription of speech. The purpose of this research was to compare different technologies for recording speech samples. Comparisons were made regarding the acoustic quality of the recorded speech samples as well as their quality as perceived by listeners.

Literature Review

Phonetic Transcription of Speech

Transcription is a tool that is used in the identification and analysis of speech sound errors and phonological processes (Louko & Edwards, 2001). The importance of transcription lies in the results that the process reveals. Treatment plans and goals are guided by the types of errors discovered through the transcription process. This makes accuracy essential. Speech-language pathologists break down a speech sample into individual sounds and record them phonetically. Small (2016, p. 413) defined phonetics as “the study of the speech sounds, their acoustic and perceptual characteristics, and how they are produced by the speech organs.” Phonetic transcription involves using a special alphabet designed to represent the actual speech sounds in words rather than their traditional spellings. Most speech-language pathologists use the International Phonetic Alphabet for this purpose (International Phonetic Association, 1999). There are two types of transcription: broad transcription and narrow transcription. Narrow transcription contains detailed variations of each sound while broad transcription includes the basic consonants and vowels (Louko & Edwards, 2001). Narrow transcription is desired as it does not assume correct production of the sound (Louko & Edwards, 2001). Transcribing speech can be difficult to do “live,” as the speaker may speak quickly with many errors. It is desirable to record the speech so that the clinician may go back and listen to the sample as many times as needed to get an accurate and reliable transcription (Louko & Edwards, 2001). Some sounds are especially prone to transcription error, making a recording all the more beneficial.

Portability and Accessibility of Devices

Given that working with recorded speech samples is the preferred way to phonetically transcribe a speech sample, professionals in the field of speech-language pathology need equipment that will provide high quality recordings. Not all speech-language pathologists have access to dedicated digital audio recorders, but most have access to more readily available technology such as laptop computers or iPad devices. Recognizing this need, researchers have begun to investigate utilizing these types of technologies in the field of speech-language pathology, and this line of research will perhaps support the viability of using more readily available technologies for acquiring speech samples (Lin et al., 2012; Vogel & Maruff, 2008; Vogel et al., 2014).

Microphone Selection

Acquiring a high quality speech sample depends largely on the selection of a microphone and not just on the device used to make the recording. There are numerous styles and types of microphones, but certain characteristics better lend themselves to speech sample acquisition. According to Howard and Murphy (2008), factors that affect the quality of a microphone include the frequency response, the distortion rate, and the signal-to-noise ratio. An ideal microphone, although practically impossible to achieve, would transparently convert an acoustic signal to an electrical signal without altering the quality (Howard & Murphy, 2008). For the purposes of speech and language assessments as well as research in the field of speech-language pathology, we want a microphone that focuses solely on the speaker, with no other interfering noise. Noise levels of microphones refer to the output level of the microphone when isolated from all other sounds (Howard & Murphy, 2008). Selecting a microphone with minimum noise levels is important so as to not distort the speech sample.

Various technologies such as laptop computers have internal microphones within their systems. Vogel and Maruff (2008), however, did not recommend using such a microphone for speech research due to the lower grade of technology the system presents. Additionally, controlling the distance between the speaker's mouth and the microphone would be more difficult with built in microphones. USB-based microphones might be viable alternatives due to their ability to bypass the soundcard device of the laptop. With a USB microphone, analog signals are converted to digital signals before they even reach the computer (Vogel & Maruff, 2008). USB-based microphones offer an affordable and simple alternative as they eliminate the need for traditional acquisition systems. Vogel and Maruff (2008) discovered that a laptop equipped with a USB-based microphone was able to produce results on certain acoustic measures that were comparable to higher-quality systems. This suggested the potential for successful application of cost-effective and readily available technology in the acquisition of speech samples.

Deliyski, Shaw, and Evans (2005), as well as Barsties and De Bodt (2014), suggested condenser-type microphones were the best to use for voice recordings. These authors cited a better focus on the voice signal and less of a focus on the background noise. The levels of noise in the acoustic environment must be carefully monitored, as noise levels could affect the signal. Deliyski et al. recommended the noise levels be below 30 decibels in the recording environment. Acceptable signal-to-noise ratio values range from 30 to 42 decibels (Barsties & De Bodt, 2014). The frequency range of the microphone is also important, as it should be able to catch the entirety of the spectrum of the human voice: 20 to 20,000 hertz (Barsties & De Bodt, 2014). The distance of the speaker from the microphone should remain constant throughout, making a head-mounted microphone a favorable option (Barsties & De Bodt, 2014).

Listener Perception

Because transcription is a perceptual task, Howard and Murphy (2008, p. 95) proposed that "the most expensive microphone is not always the best one for the job." The acoustic analysis data might indicate one thing, but if listeners do not perceive the signal as having an appropriate standard of quality, the sample would be of little help in assessing speech. Listener perception is not always reliable, however, due to the introduction of errors by the listeners such as memory and attention deficits, fatigue, and the potential for mistakes (Barsties & De Bodt, 2014). Perception might be influenced by the type of microphone chosen for the particular task because the signal quality could vary by introducing warm, harsh, or even muffled qualities to the sound (Howard & Murphy, 2008).

The previous research regarding the recording of voice samples pinpointed many factors to consider. These included the device used for making the recording, microphone quality, background noise considerations, and the positioning of the microphone relative to the speaker. A search of the existing literature revealed no studies that have specifically addressed acquiring a recording for phonetic transcription of speech. Thus, the present study built on prior research by focusing on methods for acquiring speech samples for the purpose of transcribing speech sounds. Specifically, the present study included both listener judgments of the quality of the recorded speech samples as well as acoustic analyses of the samples. Because phonetic transcription is a listening task, the listener judgments might provide the most valuable information for determining the quality of the recordings.

Research Questions

This research aimed to answer the following questions:

1. What combination of microphones and recording devices provides the clearest speech sample in terms of acoustic analysis?
2. Do listener preferences align with the quality of the recordings based on the acoustic analysis?

The researcher predicted that the Marantz portable solid state recorder technology combination would produce speech samples with the highest acoustic quality, but that when compared with listener perception, there would be no obvious differences among the various technology combinations. Such results would support the use of more practical and readily available technologies by speech-language pathologists in acquiring speech samples.

Methodology

Participants

Twenty-five participants falling within the range of 19-23 years of age with no known speech or hearing difficulties and a healthy status at the time of participation were recruited. Participants were recruited via fliers placed throughout the Roy Eblen Speech and Hearing Clinic and through a departmental email reaching all current undergraduate students within the communication sciences and disorders major. All participants gave their written consent to participate prior to starting the study. Examples of the participant consent forms used in this study are included in Appendix A.

All participants were students attending the University of Northern Iowa. Three of the participants were males and 22 were females. Five of the participants were used in acquiring speech samples and twenty were used in collecting listener perception data. The 5 participants who provided speech samples for the study ranged in age from 22 to 23 years old. This group included 1 male and 4 females. The 20 participants who served as listeners for the study ranged in age from 19 to 23 years old. This group included 2 males and 18 females. All participants reported that they had hearing within normal limits, were native speakers of English, and were free of any speech or language disorders at the time of the study. Additionally, all participants were able to participate in conversation and perceive speech at normal loudness levels.

Instrumentation

Recordings were made using combinations of microphones and recording devices. The combinations included a 64 gigabyte, second generation iPad equipped with a “Shure” MVL omnidirectional, condenser lavalier microphone, a Lenovo laptop (Intel® Core™ i5 3230M CPR processor at 2.60GHz, 64-bit operating system, operating with Windows 10 Home) running Audacity recording software equipped with a MOVO USB-M1 omnidirectional, condenser lavalier microphone, and a Marantz portable solid state recorder (model PMD660) equipped with an Audio-technica AT803 omnidirectional, condenser lavalier microphone.

For the listening portion of the study, the researcher presented the speech samples using the Lenovo laptop computer. Each participant listened to the speech samples through a set of Maxell, on-ear headphones provided by the researcher.

Procedures

Participants in the speech sample recording phase were asked to read “the Grandfather Passage” (see Appendix B; Reinstein, n.d.) and take part in a brief conversation following the prompt, “please describe your favorite vacation and why.” Each participant was recorded in the Roy Eblen Speech and Hearing Clinic in a quiet room. Participants were recorded using all three technology combinations simultaneously. In total, thirty speech samples were collected. Although head mounted microphones are often preferable, the researcher selected lavalier microphones for this study because this allowed for simultaneous recording of the sample speech samples by all three devices. The speakers wore a lanyard around their neck and each of the three lavalier microphones were mounted on this lanyard. The researcher mounted the microphones at a consistent microphone-to mouth distance of approximately 15 centimeters.

After speech samples were collected, the samples were cut down using an acoustic analysis software, Praat (Boersma & Weenink, 2017). The researcher created 15-30 second clips to be used for the listener perception task. The first fifteen seconds of the conversational samples were used in addition to the time necessary to end the sample following a complete sentence.

The first four sentences of “the Grandfather Passage” were used. Once each speech sample had been shortened, they were given a random number that would correspond to the order in which the listener would hear them.

Participants within the listener perception phase met one-on-one with the researcher in a quiet environment and listened to each of the thirty speech samples individually. After each speech sample ended, they were asked to complete two questions as a part of their listener questionnaire. A Likert scale was used to collect this data. A sample of the listener perception questionnaire is included in Appendix C. Participants were asked to rate the sample clarity and the presence of noise in the signal. For sample clarity, listeners had the options of *bad, poor, fair, good*, or *excellent*. For presence of noise, the choices were *noise is all I hear, noise is perceptible and distracting, noise is perceptible, noise is just perceptible,* and *noise is undetectable*. The numbers one through five were associated with each answer, with one representing the lowest quality and five representing the highest and most desirable.

Data Analysis

After listener perception ratings were completed by all 20 participants, the process of data analysis began. In analyzing the listener perception questionnaires, averages were developed for both the clarity and presence of noise ratings using the numbers associated with each Likert selection. In addition, each response was tallied to better represent the spread of selections across each technology combination.

To analyze each individual speech sample recording, the full-length files were saved as long sound files and opened using the acoustic analysis software previously mentioned; Praat

(Boersma & Weenink, 2017). The Praat software was utilized to assess the signal-to-noise ratio and frequency of amplitude clipping for all 30 speech samples.

The procedure for identifying peak clipping using the Praat software came from Gouskova (2016) and “Sound: Scale intensity..” (2012). The Praat software displayed signal values up to ±1 (100 decibels). In analyzing the speech samples, any sample that contained a point that reached beyond the 100 decibel point was recorded as having been clipped during the recording process.

Signal-to-noise ratio data were also collected using the Praat software. The maximum air pressure of the entire sound signal was collected using a Praat analysis feature. The absolute value was recorded and used as the “signal” value in the signal-to-noise ratio. To determine the noise value, each speech sample was divided into five even sections. Once the sections were determined, the first .25 second noise value that was present was selected. The definition for noise was a segment of at least .25 seconds that was free of speech. Once again, the absolute value of the maximum air pressure of this noise selection was recorded. This process took place within each of the five sections of every sample. The five noise values were then averaged, and this became the “noise” value in the signal-to-noise ratio. When all of the signal and noise values were determined, the ratios were calculated by dividing each sample’s signal value by its corresponding noise value.

After signal-to-noise ratios were calculated for all 30 speech samples, the mean and standard deviation for each speech sample was determined and a 95 percent confidence interval was created using the standard error of measurement. Results were calculated separately for the Grandfather Passage and for the conversational sample.

Reliability

An independent rater reanalyzed 20 percent of the speech samples using the Praat software. Data collected from two speakers and all three technology combinations were randomly chosen and reassessed for accuracy purposes. The independent rater found no differences from the original data collected.

Results

Analysis of Signal-to-Noise Ratio

The signal-to-noise ratios (SNRs) were determined for each speech sample and were separated according to the sample type (i.e. Grandfather Passage, conversational sample). The results for the Grandfather passage are shown in Table 1. The mean SNR for the Marantz technology combination was 21.64 (*SD* =3.35), the mean for the laptop technology combination was 46.10 (*SD* =12.48), and the mean for the iPad technology combination was 63.49 (*SD* = 22.38). The findings for the conversational speech recordings followed a similar pattern (see Table 2). The mean SNRs were 16.68 (*SD* =3.00) for the Marantz technology combination, 40.14 (*SD* = 6.56) for the laptop technology combination, and 50.24 (*SD* = 14.69) for the iPad technology combination.

Table 1. Mean SNRs from the Recordings of the Grandfather Passage

	Marantz	Laptop	iPad
Mean SNR	21.64	46.10	63.49
Standard Deviation	3.35	12.48	22.38

Table 2. Mean SNRs from the Recordings of Conversational Speech

	Marantz	Laptop	iPad
Mean SNR	16.68	40.14	50.24
Standard Deviation	3.00	6.56	14.69

Figures 1 and 2 display the mean signal-to-noise ratios and standard error of measurement (SEM) using a 95 percent confidence interval. For both the Grandfather Passage and the conversational samples, the Marantz technology combination was found to have the lowest signal-to-noise ratio, whereas the iPad technology combination consistently had the highest. In this sense, the signal-to-noise ratio data was favorable to the iPad technology combination. For both the recordings of the Grandfather Passage and the conversational speech samples, the confidence intervals for the Marantz recordings and the iPad did not overlap and indicated a clear difference in SNR.

Figure 1. Mean SNR and 95% Confidence Interval for Recordings of the Grandfather Passage

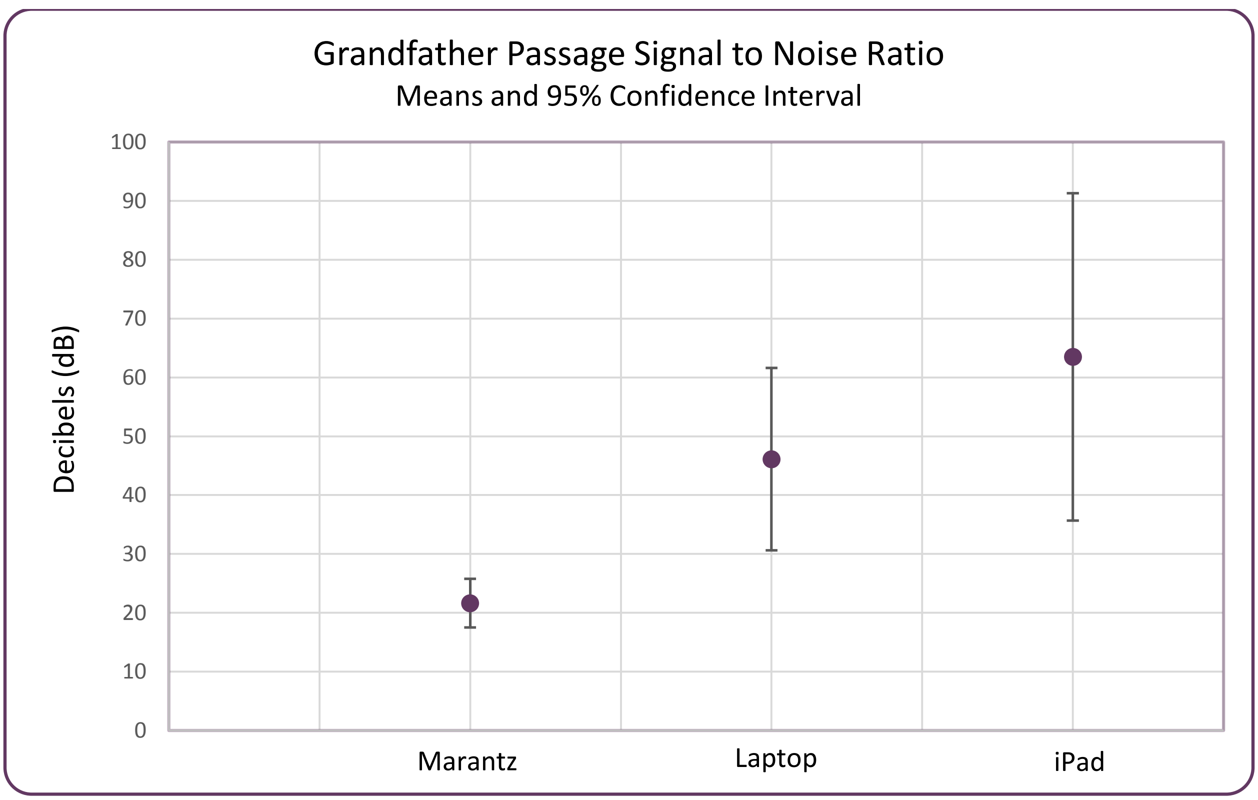
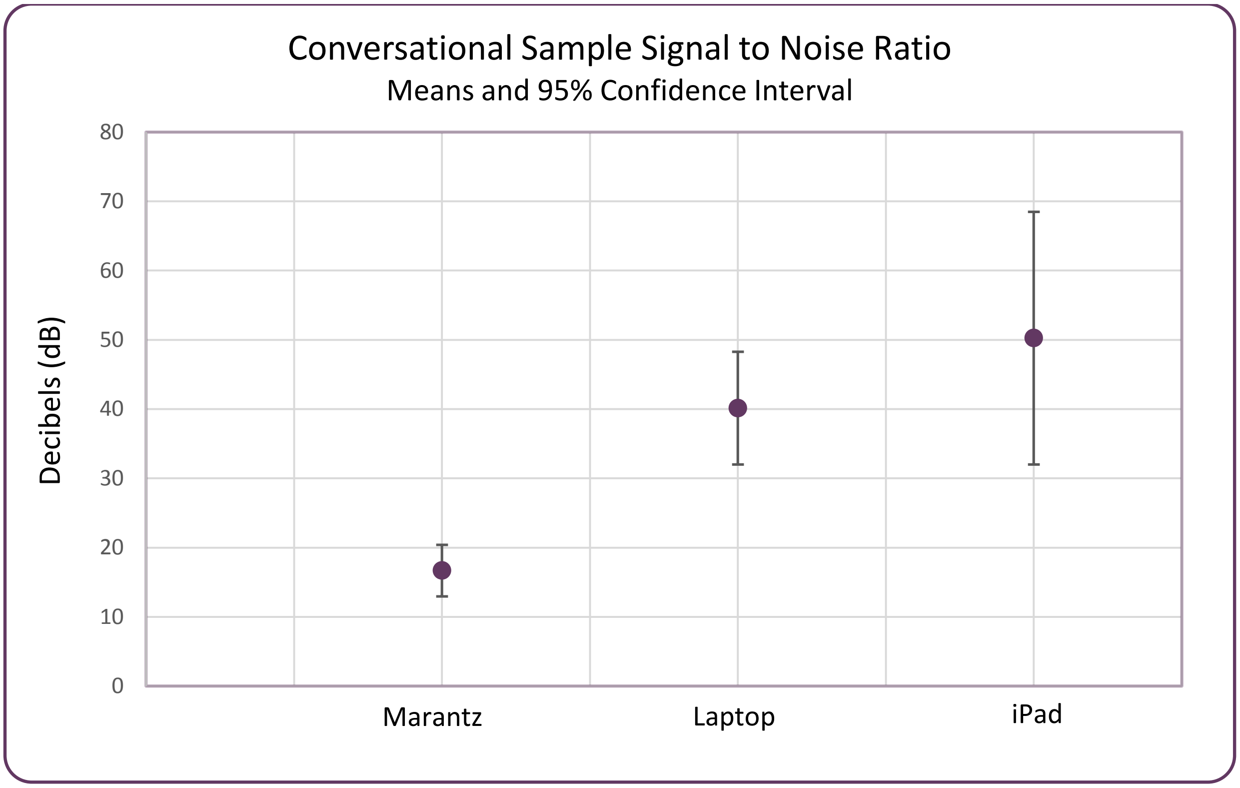


Figure 2. Mean SNR and 95% Confidence Interval for Recordings of Conversational Speech



While the Marantz technology combination recorded the lowest signal-to-noise ratio, it also had the lowest variability of the three technology combinations with the iPad technology combination having the most variability. The Audacity technology combination consistently fell in the middle of these figures. Due to the notable differences in variability, no further analysis were done using the signal-to-noise ratio data.

The researcher followed up these findings by analyzing signal and noise values separately. It is interesting to note that whereas the Marantz technology combination had the lowest signal-to-noise ratio, it did not have the lowest signal value. The average signal and noise values can be found in Table 3. The Marantz technology combination recorded the second highest signal value, just under the iPad technology combination which had the highest of the three. Noise value, however, greatly affected the Marantz technology combination's signal-to-noise ratio as it recorded the highest noise value over the other two technology combinations.

Table 3. Mean Signal and Noise Values for the Three Recording Technologies

Average Signal Values

	Marantz	Laptop	iPad
Grandfather Passage	0.530	0.173	0.535
Conversational Sample	0.413	0.162	0.416
Overall	0.472	0.168	0.476

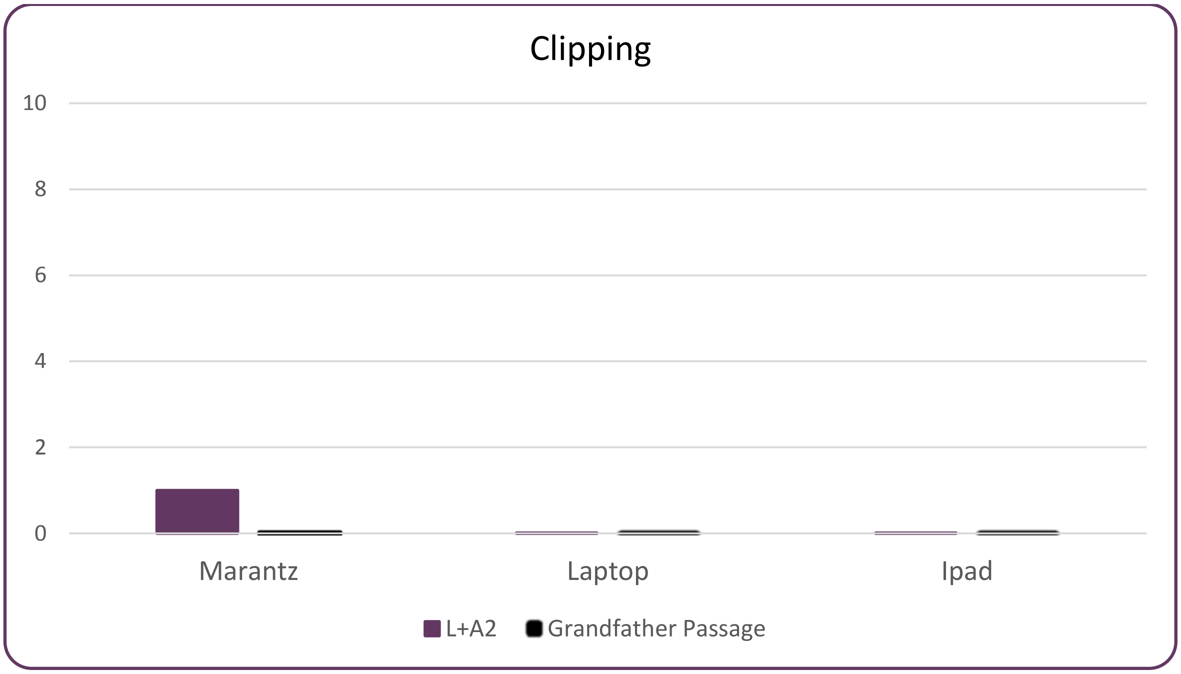
Average Noise Values

	Marantz	Laptop	iPad
Grandfather Passage	0.025	0.004	0.009
Conversational Sample	0.026	0.004	0.009
Overall	0.026	0.004	0.009

Analysis of Clipping

Figure 3 shows the incidence of peak amplitude clipping within the speech samples. Only one of the thirty speech samples was found to have clipping. The speech sample in which it occurred was recorded using the Marantz technology combination. The absence of clipping is consistent with the recording procedures used in the present study. Each recorder was deliberately set up to avoid clipping during the acquisition of the speech samples. No further analysis was done regarding clipping.

Figure 3. Instances of Peak Amplitude Clipping across all Recordings



Analysis of Listener Perception Ratings

The results from each participant's listener perception questionnaire were tallied and average ratings were calculated for both the Grandfather Passage and the conversational sample under each technology combination. The listener ratings for presence of noise are shown in Table 4 and the average ratings for presence of noise are shown in Table 5. Listener perception data revealed that listeners did not prefer one technology combination over another. For the Grandfather Passage, the listeners infrequently reported a distracting level of noise. Only 3 to 4 listeners reported “noise is perceptible and distracting” for any of the devices. The majority of listeners reported that noise was undetectable. These values were 49 for the iPad recordings, 58 for the Marantz recordings, and 60 for the laptop recordings. The mean ratings were 4.42 for the Marantz, 4.41 for laptop, and 4.41 for the iPad. The overall values fell between “noise is just perceptible” and “noise is undetectable.”

Table 4. Listener Ratings of the Presence of Noise in the Recorded Signals

Presence of Noise in the Signal- Tally Total

Grandfather Passage

	Noise is all I hear (1)	Noise is perceptible and distracting (2)	Noise is perceptible (3)	Noise is just perceptible (4)	Noise is undetectable (5)
Marantz	0	4	8	30	58
iPad	0	4	22	25	49
Laptop	0	3	13	24	60
	0	11	43	79	167

Conversational Sample

	Noise is all I hear (1)	Noise is perceptible and distracting (2)	Noise is perceptible (3)	Noise is just perceptible (4)	Noise is undetectable (5)
Marantz	0	6	30	29	35
iPad	0	14	28	34	24
Laptop	0	8	22	40	30
	0	28	80	103	89

For conversational speech, the listeners reported a distracting level of noise slightly more often. For the iPad technology combination, 14 listeners reported “noise is perceptible and distracting.” In comparison, 6 listeners rated noise as perceptible and distracting for the Marantz technology combination's recordings and 8 for the laptop technology combination's recordings. The majority of listeners reported that noise was "just perceptible" for the conversational speech recordings. The mean ratings were 3.93 for the Marantz, 3.92 for the laptop, and 3.68 for the iPad. The overall values fell between "noise is perceptible" and "noise is just perceptible" for the samples of conversational speech.

Table 5. Mean Listener Ratings of Presence of Noise in the Recorded Signals

Listener Perception Average Ratings

Presence of Noise in Signal

	Grandfather Passage	Conversational Sample	Overall
Marantz	4.42	3.93	4.175
iPad	4.19	3.68	3.935
Laptop	4.41	3.92	4.165

Only small differences among the technology combinations were found when looking at the average listener ratings for presence of noise in the various samples. The Grandfather Passage was reported as having less perceptible noise than the conversational samples across all technology combinations. This can be seen in Table 5. Due to the type of data gathered using the listener ratings, no formal statistical analysis was performed. The data did not fit with a Chi Square analysis because several of the cells had values that fell below 5 instances.

The listener ratings for clarity are shown in Table 6 and the average ratings for clarity are shown in Table 7. The listener ratings of clarity revealed that listeners did not have a clear preference for one technology combination over another. For the Grandfather Passage, the listeners occasionally rated the clarity as “bad” for all three technology combinations. Table 7 showed that 3 to 5 listeners rated clarity as “poor.” The majority of listeners reported that clarity was excellent. These values were 45 for the laptop recordings, 48 for the iPad recordings, and 52 for the Marantz recordings. The mean ratings were 4.26 for the Marantz, and 4.24 for laptop and the iPad. The overall values fell between “good” and “excellent.”

Table 6. Listener Ratings of the Clarity of the Recorded Signals

Clarity- Tally Total

Grandfather Passage

	Bad (1)	Poor (2)	Fair (3)	Good (4)	Excellent (5)
Marantz	2	5	10	31	52
iPad	1	3	15	33	48
Laptop	1	3	12	39	45
	4	11	37	103	145

Conversational Sample

	Bad (1)	Poor (2)	Fair (3)	Good (4)	Excellent (5)
Marantz	1	2	14	45	38
iPad	1	3	20	34	42
Laptop	0	4	16	48	32
	2	9	50	117	112

The listener ratings of clarity of the conversational speech samples had some similarities with the ratings of the Grandfather Passage, but also some differences. A similarity was that listeners rated the clarity of the signal as “bad” or “poor” infrequently. A rating of “bad” occurred only 0 to 1 times across the three devices and a rating of “poor” occurred only 2 to 4 times. A difference was that listeners rated the clarity as “excellent” less often for the conversational speech samples. For the conversational speech samples, 117 listeners reported that clarity was “good” and 112 reported that clarity was “excellent.” The mean ratings were 4.17 for the Marantz, 4.08 for the laptop, and 4.13 for the iPad. As with the recordings of the Grandfather Passage, the overall ratings of clarity of the conversational speech recordings fell between “good” and “excellent.” As with the presence of noise, the listeners rated sample clarity of the different devices in a similar way. The listeners did not appear to have clear preference for any particular technology combination based on the presence of noise or the clarity of the recorded signal.

Table 7. Mean Listener Ratings of the Clarity of the Recorded Signals

Listener Perception Average Ratings

Clarity

	Grandfather Passage	Conversational Sample	Overall
Marantz	4.26	4.17	4.215
iPad	4.24	4.13	4.185
Laptop	4.24	4.08	4.16

Discussion

The present study had essentially two components; acoustic analysis and listener judgments of the recorded speech samples. The basic premise was to explore if readily available technology could be utilized within the practice of speech-language pathology to acquire speech samples. As Louko and Edwards (2001) stressed, recording speech is important for accuracy and reliability in terms of transcription and the ability to replay samples as needed. Before considering the findings of this research, a few factors that might have impacted the results must be addressed.

Microphone Selection

The significance of being able to rely on readily available technology is largely due to the cost variations in recording devices. This was a limiting factor in the current research study, as price range was monitored in microphone selection to ensure cost effective microphones were used. Whereas Marantz recorders can cost up to $600, the microphones used with the laptop and iPad both came in under $80. This is a significant price difference, and so it was a goal of this research to determine, if in fact, the price made an impact on the quality; both in terms of acoustic analysis and listener perception. Previous research has found that laptops equipped with USB-based microphones were able to produce results comparable to higher-quality systems, so the researcher in the current study was confident in the microphones selected for this study (Vogel & Maruff, 2008). One issue that might have impacted the findings of this study was that the microphone selected for the Marantz recorder was not at the highest quality available. This occurred because the researchers needed to use lavalier microphones so the speakers could wear all three microphones at the same time. Microphone selection is important because quality of the recordings is based on the device/microphone combinations and not on the devices alone.

The cost of the microphones was not the only consideration when making the selections. Recordings were to be done simultaneously, therefore the researcher had to rule out head-mounted microphones. Barsties and De Bodt (2014) supported the use of head-mounted microphones due to their ability to keep the distance of the speaker from the microphone consistent, so careful attention was paid to this consistency throughout the recording process. Device mounted microphones were also explored, but it was determined that lavalier microphones would provide the best fit for the specific needs of this research. It has been suggested by multiple studies that condenser-type microphones be used in voice recordings, so the researcher in the current study used that advice while making microphone selections (Barsties & De Bodt, 2014; Deliyski et al., 2005). In addition, microphones that were highly rated for this particular intended use were chosen.

The Grandfather Passage

The Grandfather Passage was chosen for this research due to its common appearance in the field of speech-language pathology. The passage is designed to provide opportunities to pronounce most American English phonemes, which provided this study with an excellent overview of the effects that these phonemes have on both acoustic analysis and listener perception. In addition, this passage is available in the public domain.

Findings of the Acoustic Analyses

The researcher conducted acoustic analyses of the recorded speech samples as one strategy to determine what combination of microphones and recording devices provided the clearest speech samples. For the current study the researcher used two basic acoustic analyses, determination of the signal-to-noise ratios (SNRs) in the recorded samples and identification of peak amplitude clipping in the signals. The analysis yielded means and standard deviations for each type of speech task (i.e. Grandfather Passage and conversational sample). In both types of speech samples, the iPad equipped with a Shure MVL external microphone yielded the highest SNRs, whereas the Marantz recorder equipped with an Audio-technica AT803 microphone yielded the lowest SNRs. These results were contrary to our predictions, as higher signal-to-noise ratios are desirable for quality speech recordings. As Howard and Murphy (2008, p. 95) have pointed out, however, "the most expensive microphone is not always the best one for the job." The current research determined the technology combination of the second generation iPad equipped with a "SHURE" MVL omnidirectional, condenser lavalier microphone provided the clearest speech sample in terms of acoustic analysis (signal-to-noise ratio) when compared with the two other technology combinations.

The recorded samples were also analyzed to identify instances of peak amplitude clipping because the presence of clipping would result in a distorted speech signal. Peak clipping was not a factor in the present study because it occurred only once across all of the recorded samples. This meant the researcher was successful in adjusting the record levels set during the speaking tasks to avoid peak clipping.

Findings from the Listener Judgment Task

The second phase of this study involved gathering information from a listener perception task. Twenty listeners provided judgments regarding the presence of noise and the perceived clarity of the recorded speech samples. The findings revealed that listener judgments did not align with the quality of the recordings based on the acoustic analysis (i.e., signal-to-noise ratio). As discussed above, the acoustic analysis demonstrated a clear difference in the SNRs of the iPad technology combination compared to the Marantz technology combination which favored the iPad. However the listener perception data was not as distinct. The results revealed no clear differences among the three devices when judged by listeners for both presence of noise in the signal and clarity of the signal. These results suggested that signal-to-noise ratio data did not represent all that we need to know in terms of the quality of a speech sample. Further research is needed to determine why differences can exist in an acoustic analysis with no such differences appearing in the listener ratings. In answering the research question, the current data suggested that in terms of listener perception, readily available technology produced speech samples that were as acceptable as those recorded with higher-priced, professional recording devices.

Directions for Further Research

One of the limitations of the present study is that the research compared one professional quality device and microphone to two readily available devices, whereas, many professional quality recording devices are available to speech-language pathologists. Future research that includes addition professional quality devices and microphones is needed to further explore the possibility that readily available technology and microphone combinations can compete with professional quality devices in terms of both acoustic analysis and listener perceptions.

The variety of microphone and device combinations that speech-language pathologists could potentially use is extensive. Additionally, those making recordings might want to use the built in microphones in devices such as laptop computers and iPads. For this reason, it would be reasonable to continue this line research using additional microphones paired with the built-in microphones that come in various readily available devices. Along with this point, the current research used neither the most expensive nor the least expensive microphones available for the particular devices utilized, so the cost of the technology combination and its associated quality is an area that warrants further exploration to determine if an increase in cost relates to an increase in performance.

As noted above, the findings from this study left an important question unanswered. Further research is needed to determine why the results could yield a clear cut difference among the device/microphone combinations, i.e., between the iPad and the Marantz recorder based on acoustic analysis, but no such difference based on listener judgments. Additional analysis may be necessary both acoustically and perceptually to get a better idea of why such a discrepancy exists. The measure used in the present study, signal-to-noise ratio, might not reflect the qualities that are most important to listeners.

Conclusion

The current research study aimed to look at various technology combinations involving both professional-grade recording devices and those more practical and readily available. Speech samples were acquired using three different technology combinations, and analyses were done on all thirty of the collected samples. When analyzed acoustically, the iPad technology combination had a consistently higher signal-to-noise ratio compared to the Marantz technology combination. The recordings made with Audacity on a laptop computer consistently fell in the middle. In this sense, signal-to-noise ratio favored the iPad combined with the Shure microphone. When this data was compared with that of the listener perception data, a different pattern emerged. No technology combination stood out as more superior in terms of the presence of noise in the signal and the signal clarity. The findings of this research should be regarded as preliminary for the reasons discussed above. However, for transcription purposes, this research suggested that readily available, cost effective technology could generate speech samples with comparable recording quality to that of a higher priced, dedicated recording device. This opens the door for the field of speech-language pathology to rely on and utilize these technologies making the acquisition of speech samples cheaper and more accessible. This has the potential to directly benefit the clients they serve. This research may also benefit broader audiences as it may act as a catalyst in continuing to explore the potential qualities and uses for readily available technologies in all fields of study.

References

Barsties, B., & De Bodt, M., (2014). Assessment of voce quality: Current state-of-the-art. *Auris Nasus Larynx, 42*(2015), 183-188.

Boersma, P.l & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Deliyski, D. D., Shaw, H. S., & Evans, M. K., (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice, 19*(1), 15-28.

Gouskova, M. (2016, September 3). Praat tutorial. Retrieved from https://www.gouskova.com/2016/09/03/praat-tutorial/

Howard, D. M., & Murphy, D. (2008). *Voice science, acoustics, and recording.* San Diego, CA: Plural Publishing, Inc.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge, UK: Cambridge University Press.

Lin, E., Hornibrook, J., & Ormond, T., (2012). Evaluating iPhone recordings for acoustic voice assessment. *Folia Phoniatrica et Logopaedica, 64,* 122-130.

Louko, L. J., & Edwards, M. L. (2001). Issues in collecting and transcribing speech samples. *Topics in Language Disorders, 21*(4), 1-11.

Reinstein, A. (n.d.). *Grandfather Passage.* Retrieved from www.amyspeechlanguagetherapy.com/grandfather-passage.html

Small, L., H. (2016). *Fundamentals of phonetics: A practice guide for students* (4th edition). Boston, MA: Pearson.

Sound: Scale intensity … (2012). Retrieved from http://www.fon.hum.uva.nl/praat/manual/Sound__Scale_intensity___.html

Speech-language pathologists. (n.d.). Retrieved from http://www.asha.org/Students/Speech-Language-Pathologists/

Vogel, A. P., & Maruff, P. (2008). Comparison of voice acquisition methodologies in speech research. *Behavior Research Methods, 40*(4), 982-987.

Vogel, A. P., Rosen, K. M., Morgan, A. T., & Reilly, S., (2014). Comparability of modern recording devices for speech analysis: Smartphone, landline, laptop, and hard disk recorder. *Folia Phoniatrica et Logopaedica, 66,* 244-250.

Appendix A

Informed Consent Documents

UNIVERSITY OF NORTHERN IOWA
HUMAN PARTICIPANTS REVIEW
INFORMED CONSENT

Project Title: A Comparison of Technologies for Recording Speech

Name of Investigator(s): Alexa Klimes

Invitation to Participate: You are invited to participate in a research project conducted through the University of Northern Iowa. The University requires that you give your signed agreement to participate in this project. The following information is provided to help you made an informed decision about whether or not to participate.

Nature and Purpose: The purpose of this research is to compare different technologies for recording speech samples for transcription purposes. Comparisons will be made regarding their acoustic quality as well as their quality as perceived by participants.

Explanation of Procedures: Each participant will listen to thirty auditory samples in the Roy Eblen Speech and Hearing Clinic and provide judgements of each sample via questionnaire of their perceived quality using parameters such as clarity and presence of noise. This data will then be compared with acoustic analysis data. Completion of this research should take each participant roughly thirty minutes.

Discomfort and Risks: There are no foreseeable risks to participation.

Benefits and Compensation: Individual participants will receive no direct benefits or compensation.

Confidentiality: A code number will be assigned to each participant for data collection and analysis. Only the research team will have access to participant identification information from the signed informed consent form. The summarized findings may be published in an academic journal or presented at a scholarly conference with no direct identifiers being revealed.

Right to Refuse or Withdraw: Your participation is completely voluntary. You are free to withdraw from participation at any time or to choose not to participate at all, and by doing so, you will not be penalized or lose benefits to which you are otherwise entitled.

Questions: If you have questions about the study you may contact Alexa Klimes at 319-651-5156 or the project investigator's faculty advisor Dr. Lauren Nelson at the Department of Communication Sciences and Disorders, University of Northern Iowa 319-273-6806. You can also contact the office of the IRB Administrator, University of Northern Iowa, at 319-273-6148, for answers to questions about rights of research participants and the participant review process.

Agreement:

> **I am fully aware of the nature and extent of my participation in this project as stated above and the possible risks arising from it. I hereby agree to participate in this project. I acknowledge that I have received a copy of this consent statement. I am 18 years of age or older.**

______________________________ __________________
(Signature of participant) (Date)

____
(Printed name of participant)

______________________________ __________________
(Signature of investigator) (Date)

______________________________ __________________
(Signature of instructor/advisor) (Date)

UNIVERSITY OF NORTHERN IOWA
HUMAN PARTICIPANTS REVIEW
INFORMED CONSENT

Project Title: A Comparison of Technologies for Recording Speech

Name of Investigator(s): Alexa Klimes

Invitation to Participate: You are invited to participate in a research project conducted through the University of Northern Iowa. The University requires that you give your signed agreement to participate in this project. The following information is provided to help you made an informed decision about whether or not to participate.

Nature and Purpose: The purpose of this research is to compare different technologies for recording speech samples for transcription purposes. Comparisons will be made regarding their acoustic quality as well as their quality as perceived by participants.

Explanation of Procedures: This study will involve obtaining audio recordings of adult participants. Participants will be asked to read a standard passage and word list and produce a conversational speech sample on a specified topic. Recordings will be taken in the Roy Eblen Speech and Hearing Clinic. A variety of recording device and microphone combinations will be used including a dedicated digital audio recorder, an iPad, and a personal computer. The recorded samples will be analyzed using acoustic analysis software (PRAAT) to assess the signal to noise ratio and high frequency clipping. The recording session should last no more than thirty minutes.

Discomfort and Risks: There are no foreseeable risks to participation.

Benefits and Compensation: Individual participants will receive no direct benefits or compensation.

Confidentiality: A code number will be assigned to each participant for data collection and analysis. Only the research team will have access to participant identification information from the signed informed consent form. Audio recordings will be stored on a USB drive secured with a password. All data will be stored in a secure area of the Roy Eblen Speech and Hearing Clinic. The summarized findings may be published in an academic journal or presented at a scholarly conference with no direct identifiers being revealed. The recorded speech samples may be used in future instructional activities with graduate and undergraduate students learning how to phonetically transcribe speech or in future studies with no direct identifiers being revealed.

Right to Refuse or Withdraw: Your participation is completely voluntary. You are free to withdraw from participation at any time or to choose not to participate at all, and by doing so, you will not be penalized or lose benefits to which you are otherwise entitled.

Questions: If you have questions about the study you may contact Alexa Klimes at 319-651-5156 or the project investigator's faculty advisor Dr. Lauren Nelson at the Department of

Communication Sciences and Disorders, University of Northern Iowa 319-273-6806. You can also contact the office of the IRB Administrator, University of Northern Iowa, at 319-273-6148, for answers to questions about rights of research participants and the participant review process.

Agreement:

> **I am fully aware of the nature and extent of my participation in this project as stated above and the possible risks arising from it. I hereby agree to participate in this project. I acknowledge that I have received a copy of this consent statement. I am 18 years of age or older.**

________________________________ ____________________
(Signature of participant) (Date)

____
(Printed name of participant)

________________________________ ____________________
(Signature of investigator) (Date)

________________________________ ____________________
(Signature of instructor/advisor) (Date)

Appendix B

The Grandfather Passage

Grandfather Passage

You wish to know all about my grandfather. Well, he is nearly 93 years old, yet he still thinks as swiftly as ever. He dresses himself in an old black frock coat, usually several buttons missing. A long beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. When he speaks, his voice is just a bit cracked and quivers a bit. Twice each day he plays skillfully and with zest upon a small organ. Except in the winter when the snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, “Banana oil!” Grandfather likes to be modern in his language.

Appendix C

Listener Perception Questionnaire Sample

Sample #1

Sample Clarity:	Presence of Noise in Signal:
1 (bad)	1 (noise is all I hear)
2 (poor)	2 (noise is perceptible and distracting)
3 (fair)	3 (noise is perceptible)
4 (good)	4 (noise is just perceptible)
5 (excellent)	5 (noise is undetectable)