2017

A comparison of technologies for recording speech and the effects of speaker age

Jacquelyn Marie Knustrom
University of Northern Iowa

A COMPARISON OF TECHNOLOGIES FOR RECORDING SPEECH AND THE EFFECTS OF SPEAKER AGE

A Thesis Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors

Jacquelyn Marie Knustrom

University of Northern Iowa

May 2017

A COMPARISON OF TECHNOLOGIES FOR RECORDING SPEECH

This Study by: Jacquelyn Knustrom

Entitled: A Comparison of Technologies for Recording Speech and the Effects of Speaker Age

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

_____	_____
Date	Dr. Lauren Nelson, Honors Thesis Advisor, Communication Sciences and Disorders
_____	_____
Date	Dr. Jessica Moon, Director, University Honors Program

Abstract

The purpose of this study was to compare recordings of speech samples obtained with a dedicated recording device to those from more readily available devices. Speech recordings are used by professionals in the field of communication sciences and disorders to identify and transcribe features of an individual's speech, regardless of his or her age. This process requires high-quality recordings, but devices intended to produce such recordings are often expensive, not easily accessible, and have few uses. This research addressed the following questions: (A) What combination of microphones and recording devices provides the clearest speech sample based on a signal-to- noise ratio? (B) Does peak clipping distort recorded speech samples for certain combinations of microphones and recording devices? (C) Are differences present between the quality of recordings for different devices dependent on the subject's age (i.e., child or adult)? Participants included 4 male and 5 female adults, ages 18-26, and 4 male and 4 female children, ages 5-10. Participants identified English as their first language, had hearing within normal limits, and had no communication disorders. Each participant was recorded performing three speech tasks. Speech was recorded simultaneously on a dedicated recording device, a personal computer, and an iPad, each paired with an omnidirectional, condenser lavalier microphone. Recordings were analyzed for signal-to-noise ratio (SNR) and instances of peak clipping. Results were compared by device and age group. Analyses revealed that the iPad yielded highest average SNR, but had the most variable values in most comparisons. The dedicated recording device generally had the lowest average SNR, but was the least variable. Peak clipping was not a significant factor for any device. No adult participants produced clipped signals, while three children did. Results suggest that readily available devices such as personal computers and iPads are appropriate for the types of acoustic analysis performed in this study.

Introduction

This project investigated the quality of recordings obtained with a dedicated recording device in comparison to more readily available devices. High quality recordings of speech samples are essential to effective instruction and research in the field of communication disorders. Professionals in the field of communication disorders need to identify and transcribe specific details about individuals' speech. The dedicated recording devices that are intended to provide high quality signals are not readily available and can be expensive. If recordings from non-professional devices were sufficient they could be more cost effective and available in more environments.

Speech samples were recorded on three different devices in combination with three microphones. A free software application, Praat (http://www.fon.hum.uva.nl/praat/), which allows analysis, synthesis, and manipulation of speech recordings was used to determine signal-to-noise ratio and determine the presence or absence of peak clipping in recordings of three types of speech samples. Signal-to-noise ratio (SNR) is a measure which compares the strength of a signal to that of the ambient or background noise. Peak clipping occurs when the amplitude of a sound wave exceeds the range which the equipment can handle. These analyses provided comparable measures of the quality of the recordings made with different device and microphone combinations.

Review of Literature

Speech-language pathologists (SLPs) assess and treat many aspects of communication and swallowing. Included in communication are speech production and fluency, voice, resonance, language, cognition, and hearing (American Speech-Language-Hearing Association [ASHA], 2016). In order to optimize their patients' ability to communicate, SLPs must be able to

correctly diagnose and efficiently treat communication disorders. Extensive education is needed to acquire the skills to do this.

Transcription

Phonetics is an area of study that is important for the field of communication disorders. Phonetics is the scientific study of speech sounds, regardless of languages in which they are used (Louko & Edwards, 2001; Small, 2016). The sounds of a language, or phonemes, are classified by their place of production, manner of production, which refers to how the articulators are configured to achieve the sound, and whether the sound is voiced or unvoiced, referring to vocal fold vibration (Small, 2016). Speech samples can be phonetically transcribed for analysis using a specially designed system, the International Phonetic Alphabet (International Phonetic Association, 1999). Analyses of phonetic transcriptions can help SLPs make a diagnosis of a speech disorder; however, phonetic transcription of speech presents several difficulties (Louko & Edwards, 2001).

Clinicians encounter difficulties when trying to transcribe at the same rate as the client's speech. In addition, holding the attention of certain client populations and transcribing their speech simultaneously is challenging. This means that clinicians often record speech samples and transcribe them after meeting with their clients. When transcribing predetermined passages or lists, the clinician is aware of the target sounds. This can result in close listening for each phoneme, which is beneficial, or it can lead to expectation errors in which the clinician perceives a correct sound that was not produced, which is not beneficial. In contrast, conversational speech samples can be problematic because intelligibility may be low and the clinician may have limited knowledge of the target sounds. Correctness is difficult to calculate if the target word is unknown (Louko & Edwards, 2001). Although transcribing from a recorded sample cannot eliminate these

challenges entirely, transcribing from a recording has many advantages. Recordings allow replaying and segmenting shorter sections of the speech sample. When listening to the recording, transcribing is the only focus of the clinician. Additionally, the clinician can create more ideal listening conditions such as listening in a quiet setting and listening through headphones. To achieve the best results and ensure accurate transcription, the clinician needs a high quality recording (Huckvale, 2009).

Equipment Demands

When selecting equipment for recording speech, one must consider both the microphone used to obtain the sample and the recording device. The purpose of microphones in voice and speech research is to convert sound pressure signals to electric signals with characteristics that are similar to speech (Svec & Grandqvist, 2010). Because this differs from the function of most microphones, professionals who record speech need to take special considerations.

Omnidirectional microphones have equal sensitivity to sound sources from any direction. Directional microphones respond differently to different directions. Howard and Murphy (2007) stated that omnidirectional condenser microphones are best for single source sounds recorded for research purposes. A critical distance from the source, in this case the speaker, also needs to be determined. This type of microphone can be connected to a variety of recording devices which may or may not result in the same quality of recording. Also, voice characteristics differ by population which may impact the recording (Howard & Murphy, 2007).

The average fundamental frequency for an adult male is 116.5 Hz. For adult females it is 207.7 Hz, while children average 277.2 Hz (Howard & Murphy, 2007). Microphones pick up a limited range of frequencies. To produce quality recordings this range must be inclusive of all frequencies that speakers of any age and gender could produce. Internal noise of the microphone

also should be lower than even the softest possible phonations in anticipation of soft-speaking clients (Svec & Grandqvist, 2010). Clinicians should also take into account that children may have trouble sitting still and maintaining the optimal distance from the microphone. Headset microphones ensure a constant distance and reduce the effect of environmental noise (Barsties & De Bodt, 2014). In order to accurately compare the quality of recording technologies, optimal microphones must be paired with the recording devices.

Little research has been done toward determining minimal hardware requirements for quality speech recordings. However, some varying conclusions have emerged in the current literature. Vogel and Maruff (2008) analyzed identical speech samples recorded using three different methods: a dedicated recorder, a disc recorder, and a laptop computer. Each was used with a different head-mounted microphone. This study showed that specialized hardware was necessary for analysis of certain acoustic features, whereas the other methods were sufficient for investigating key features of frequency and timing. Vogel and Maruff concluded that dedicated recorders were best for recording and analyzing pitch and intensity perturbation. Other methods such as a laptop computer were sufficient for investigating characteristics such as jitter and noise-to-harmonic ratio.

Through their investigation of microphone characteristics, Svec and Grandqvist (2010) were able to compile a list of recommendations for selecting a microphone. They stated that "different phonation tasks pose different demands on microphones" (p. 365). Inexpensive microphones may be appropriate in certain situations, but if specifications including optimal distance, dynamic limits inclusive of all voice levels, and frequency limits suitable for all speakers are not available, they should not be used for voice and speech research.

Deliyski, Evans, and Shaw (2004) used multiple software systems to analyze acoustic signals recorded on several combinations of computers and microphones. Their conclusions supported the National Center for Voice and Speech's recommendation of professional, head-mounted, condenser microphones. They asserted that data acquisition systems that cannot accept this type of microphone should not be used for acoustic assessment and research. This recommendation was also supported by Barsties and De Bodt (2014), though they indicated a need for more research to document the recommendation.

Speech-language pathologists perform many analyses that require high quality speech and voice recordings. These range from phonetic transcriptions of the speech sounds in a sample to analysis of the acoustic properties of the sample, even to the level of analyzing the properties of individual speech sounds (Small, 2016). To date SLPs have limited guidance regarding the selection of microphones and recording devices for different recording purposes. The guidance that is available often is somewhat dated and does not reflect the most current recording technologies available with laptop computers or tablets and smart phones. Thus, the purpose of this research is to compare different types of recording technology and provide SLPs with information they need to select technology for recording speech samples.

Research Questions

This research will address the following questions: (A) What combination of microphones and recording devices provides the clearest speech sample based on a signal-to-noise ratio? (B) Does peak clipping distort recorded speech samples for certain combinations of microphones and recording devices? (C) Are differences present between the quality of recordings for different devices dependent on the subject's age (i.e., child or adult)?

Hypothesis

The researcher hypothesized that differences in signal-to-noise ratio would not be statistically significant. With careful selection and pairing of recording technologies and microphones, little to no clipping was expected. These results would confirm that more readily available devices are sufficient for obtaining quality speech samples.

Methodology

Participants

Participants in this study included nine adults, five females and four males, and eight children, four females and four males. One adult participant was excluded randomly from comparative analysis to keep sample sizes equal. Adult participants' ages ranged from 18 to 26 and child participants were between the ages of 5 and 10. Institutional Review Board approval was obtained and its protocols were followed for the use of human participants. Participants were recruited via email and fliers posted around the University of Northern Iowa's campus and in the department of Communication Sciences and Disorders. Recruited adults received a form describing the study and provided signatures of agreement to participate (see Appendix A). Parents of child participants signed a permission form and oral consent to participate was obtained from the children (see Appendix B). Participants were asked to identify their date of birth, gender, whether or not they have had a hearing screening and the status of their hearing, their native language, history of speech therapy, and health at time of testing. Parents of child participants were asked to provide this information. Individuals who had hearing that was not within normal limits, did not identify English as their first language, or had a communication disorder were excluded from participation. Present health was addressed to ensure health factors were not likely to alter the quality of the speech recording. If interested individuals met the

requirements, succeeding recording sessions involved three speech tasks being recorded by three distinct microphone and device combinations.

Equipment

The three devices used to record speech samples were the Marantz PMD 660, an HP 2000-2b19WM personal computer with the recording application Audacity, version 2.1.2 http://www.audacityteam.org/, and an iPad generation four with recording app Voice Record Pro. The Marantz PMD 660 represented a professional, dedicated recording device, whereas the personal computer and iPad represented more readily available technologies that are capable of obtaining speech recordings through free applications. After investigating the best microphones for recording speech, the researcher decided that each device would be paired with an omnidirectional, condenser lavalier microphone, consistent with Howard and Murphy's recommendations. Lavalier microphones feature clips for hands free use. An Audio-technia AT803 was paired with the Marantz device, a MOVO USB-MI with the personal computer, and a Shure MVL with the iPad. These microphones were chosen because they fit the determined criteria, had input compatible with their respective devices, fell within the same price range (i.e., $100 to $150), and were rated similarly in quality. The researcher selected lavalier microphones rather than head mounted microphones because of the need to record to three different recording devices simultaneously. This was not possible with a head mounted microphone. The titles of the microphone, app, and recording device combinations will be shortened to Audacity, iPad, and Marantz in this paper, though these refer to the entire recording system. Each device and microphone pairing recorded three samples for each participant.

Speaking Tasks

The researcher asked participants to produce three samples: a standard reading passage, a word list, and a conversational speech sample. Three tasks were decided on to present a variety of phonation tasks, challenging the microphones in different ways as described by Svec and Grandqvist (2010). These types of speaking tasks also represented the variety of speech samples that SLPs need to record when working with clients. The Grandfather Passage, a standard, phonetically balanced passage used in speech sample acquisition was read by adult participants. Children were asked to read a more age-appropriate phonetically balanced passage, the Caterpillar Passage. A list of 94 words was compiled to feature each English phoneme in every position: initial, medial, and final. The passages and word list are featured in Appendixes C-E. The variation of speech sounds and their positions in both the passage and the word list ensured that recording technologies could capture accurately sounds of any frequency and intensity. Participants were instructed to read the passage at a comfortable rate. When producing the word list, participants were asked to pause two to three seconds between each word on the list. For children who were not yet able to read, the investigator read one sentence or phrase of the Caterpillar Passage at a time and asked the child to repeat the sentence until the child produced the full passage. Similarly, the investigator read one word from the list at a time and the child repeated those words to complete the list. For the final speech task, participants engaged in a conversation with the investigator. They were prompted to describe their favorite vacation. The target length of the conversational sample was two minutes. If participants did not speak this long, they were further prompted with questions about their experience, such as “Who did you go with?” and “What was your favorite activity while you were there?”

Environment

Of the 18 recording sessions, 16 took place in a therapy room of a university, speech and hearing clinic. The remaining two were completed in a small, quiet room intended to have acoustic properties similar to those of the therapy rooms. Participants read or repeated first the Grandfather or Caterpillar passage, then the word list, and ended with the conversational speech sample. The speaking tasks were recorded concurrently on the three devices to minimize participant inconvenience and to provide similar samples for the devices to record. The three microphones were attached to a lanyard around the participant's neck at a measured distance of 15 centimeters from mouth to microphone. Participants were instructed to maintain this distance to the best of their ability and avoid touching the microphones with their hands or the table surface in front of them.

Analysis

All three devices produced recordings with the .wav file type. At the completion of a recording session, the recordings from all three devices were transferred to the researcher's laptop computer. All recordings were opened in the Praat acoustic analysis software. Praat is a flexible tool with many different choices for analyzing speech and other acoustic samples (Boersma & Weenink, 2017; van Lieshout, 2003). For this study, signal-to-noise ratio and instances of clipping were determined from the provided waveform view and spectrogram. Signal-to-noise ratio is a measure which compares the strength of a signal to that of the ambient or background noise. The point of highest signal intensity in each sample was used as the signal value. The noise value was computed by segmenting the recording into five sections and averaging an instance of noise from each. The duration of the signal for the passage and conversational speech sample tasks were divided by five and the first available 0.25 seconds of

noise from each segment were used for averaging. The 94 words on the list were divided into five sections and a word was randomly selected from each section. The first available 0.25 seconds of noise following the designated words were used for averaging in each of the recordings. Noise was considered to be sound produced by the recording device plus ambient or background noise. Unintentional, non-speech sounds from the speaker or environment were not included as signal or noise. The most intense signal value was divided by the average of the five instances of noise to determine signal-to-noise ratio. The signal-to-noise ratios were then averaged by participant group for each speech task and further analyzed.

Mean and standard deviation of SNR was first computed for the overall group, adults and children, and by recording equipment for each of the three tasks. The overall means were compared across the devices within tasks as an indication of overall quality of recording, as measured by SNR. Adults and children were then separated within tasks and means and standard deviations were calculated separately for each device. From these values, 95% confidence intervals were determined. The confidence intervals reflected the standard error around the mean. These values were plotted to provide side-by-side comparison of ranges for each device and microphone combination. Following analysis of SNR values, recordings were inspected for instances of clipping. The researcher originally planned to report SNR and clipping separately for the adult male and female speakers. The number of participants in each group was small and further dividing the adult group into males and females was not practical.

Clipping occurs when the amplitude of a sound wave exceeds the range which the equipment can handle. This may result in a distorted or less clear signal. As a measure to prevent clipping, sound tests were conducted prior to recording of the three speech tasks. Participants were asked to count to ten and read aloud the first two sentences of the passage. The volume unit

meter on each of the devices was monitiored during these tests and adjustments were made if the signal was entering the yellow or red area of the meter. A display of yellow or red during recording was not necessarily indicative of an occurrence of clipping. Instances in which the signal exceeded the analysis window in Praat, resulting in a value display of 1 were considered clipped (Boersma & Weenink, 2017). The number of instances of clipping was counted for each recording. Recordings in which an unintentional noise, such as a slamming door, resulted in a signal of 1 were not counted as clipping. The number of instances of clipping was compared by device and microphone combination and by sample population.

Reliability

An independent rater reanalyzed 100 of the data points for signal-to-noise ratio selected at random. The interrater percentage of agreement for this reanalysis was 83%. The Praat displays intensity values as ± for recorded sounds and maximum intensity of a signal is the absolute value. The independent rater noted that the examiner recorded the positive value 17 times when the negative value actually reflected maximum intensity. However, this difference had minimal impact on the findings of the study. Values were calculated to three decimal points and the difference was greater than .01 in only one instance. The interrater percentage of agreement using this more lenient criteria was 99%.

Results

Signal-to-Noise Ratio

Passage. The overall means and standard deviations for the recordings of the reading passage are shown in Table 1. The overall mean of SNR for the recordings captured by the MOVO microphone and Audacity personal computer app for the reading passage task was 56.176 (*SD* = 41.810). For the Shure microphone with the Voice Record Pro iPad app the mean

was 75.117 (*SD* = 46.327). The recordings from the Audio-technia microphone and Marantz recorder had a mean of 32.977 (*SD* = 22.362) for this task. These values are expressed in Table 1. The average signal-to-noise ratio was greatest for the iPad recordings, but its values were also the most variable. The Marantz produced the lowest average SNR, but was the most consistent with the smallest standard deviation. These results were consistent when data were divided into adult and child sample populations.

Table 1: Overall SNR Results for Recordings of the Reading Passages for Children and Adults

	Audacity	**iPad**	**Marantz**
Mean	56.176	75.117	32.977
Standard Deviation	41.810	46.327	22.362

Table 2 displays the means and standard deviations for SNR computed separately for the adult and child groups. These findings are also displayed in Figure 1. The adult participants' recordings from the PC had a mean SNR of 49.784 (*SD* = 19.771). This resulted in a confidence interval of 35.51-83.05 and standard error of measurement (*SEM)* of 14.27. As shown in Table 3, the recordings of child speakers from the same device had a mean of 61.235 (*SD* = 56.61), resulting in a confidence interval of 13.9-108.58 and a SEM of 47.34. The mean for adult recordings from the iPad was 65.961 (*SD* = 21.013), resulting in a confidence interval of 55.61-88.92 and a SEM of 10.35. The child recordings from the iPad had a mean of 84.274 (*SD* = 62.975), resulting in a confidence interval of 31.62-136.93 and standard error of 84.274. The Marantz recorder produced a mean of 33.702 (*SD* = 22.480) for adults. These recordings had a confidence interval of 55.61-88.92 and a SEM of 16.95. For the children, the mean of Marantz

recordings in this event was 32.777 (*SD* = 24.512), resulting in a confidence interval of 12.28-53.27.

Table 2: SNR for Adults Recorded While Reading the Grandfather Passage

	Audacity	**iPad**	**Marantz**
Mean	49.783	65.961	33.702
Standard Deviation	19.771	21.013	22.480
95% Confidence Interval	35.51-83.05	55.61-88.92	16.75-66.21
Standard Error	14.27	10.35	16.95

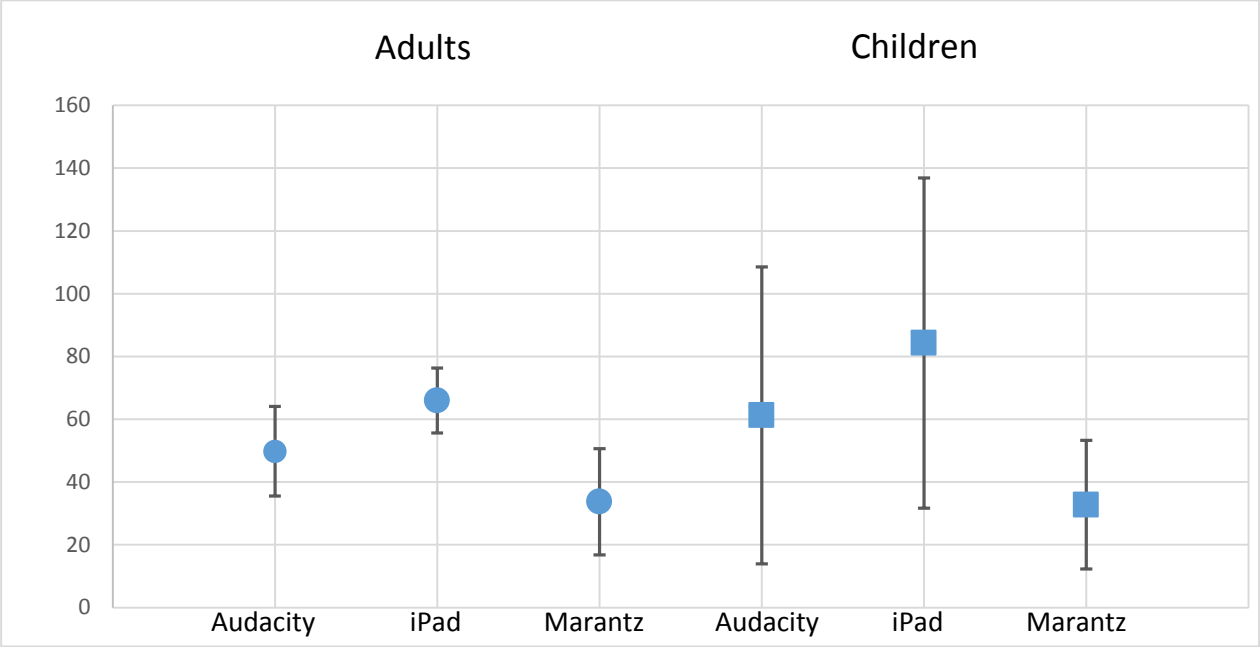
Table 3: SNR for Children Recorded While Reading the Caterpillar Passage

	Audacity	**iPad**	**Marantz**
Mean	61.236	84.274	32.777
Standard Deviation	56.616	62.975	24.512
95% Confidence Interval	13.9-108.58	31.62-136.93	12.28-53.27
Standard Error	47.34	52.65	20.5

Consistent with the overall values, the mean SNR was highest for recordings taken by the iPad for both adults and children. The standard deviation was lowest for the Marantz in the child sample, but Audacity with the personal computer produced the lowest standard deviation for adult recordings. Overall, the child recordings had higher average SNRs across all three recording apparatuses, and their values were much more variable resulting in higher standard deviations for each device. The values for adult and children in the passage task are plotted in Figure 1. As shown in this figure, the confidence intervals for all of the recording options overlapped for the child participants, meaning there was no clear advantage for any of the

device/microphone combinations. For the adults, the confidence intervals overlapped for most of the comparisons. However, recordings made with the iPad did have a better SNR than recordings made with the Marantz recorder with no overlap of the 95% confidence intervals.

Figure 1: Means and 95% Confidence Intervals for SNR from the Reading Passages



Word List. As shown in Table 4, the mean SNR of all word list recordings taken with Audacity was 70.582 (*SD* = 53.213). The iPad recordings of the word list had a mean of 120.979 (*SD* = 66.194). Word list recordings from the Marantz had a mean of 28.486 (*SD* = 21.056). The iPad recordings produced the highest average SNR and were the most variable for this speech task as well. The Marantz once again had the lowest SNR, but yielded the least variability.

Table 4. Overall SNR Results for Recordings of the Word List for Children and Adults

	Audacity	**iPad**	**Marantz**
Mean	70.582	120.979	28.486
Standard Deviation	53.213	66.194	21.056

As shown in Tables 5 and 6, the researcher also calculated the results for the adult and child groups separately. This analysis revealed the mean SNR for adult speakers for the word list recorded by Audacity was 100.307 (*SD* = 58.813), resulting in a confidence interval of 51.13-149.48 and a SEM of 49.18. For this task on the same device, the children had a mean SNR of 40.857 (*SD* = 24.272), resulting in a confidence interval of 20.56-61.15. The iPad word list recordings of adults had a mean of 166.193 (*SD* = 62.2). The confidence interval was 114.18-218.2 and SEM was 52.01. For the children, iPad recordings had a mean SNR of 75.765 (*SD* = 29.118), resulting in a confidence interval of 51.33-100.02 and a SEM of 75.765. The Marantz recordings of the word list had a mean of 38.436 (*SD* = 23.597) for adult speakers, resulting in a confidence interval of 18.7-58.17 and a standard error of 19.74. The child speaker recordings on the Marantz had a mean of 18.535 (*SD* = 12.918), resulting in a confidence interval of 7.73-29.34 and a SEM of 10.81. The data for SNR analysis of the word list recordings is presented in Tables 5 and 6 and displayed in Figure 2.

Table 5. SNR for Adults Recorded While Reading the Word List

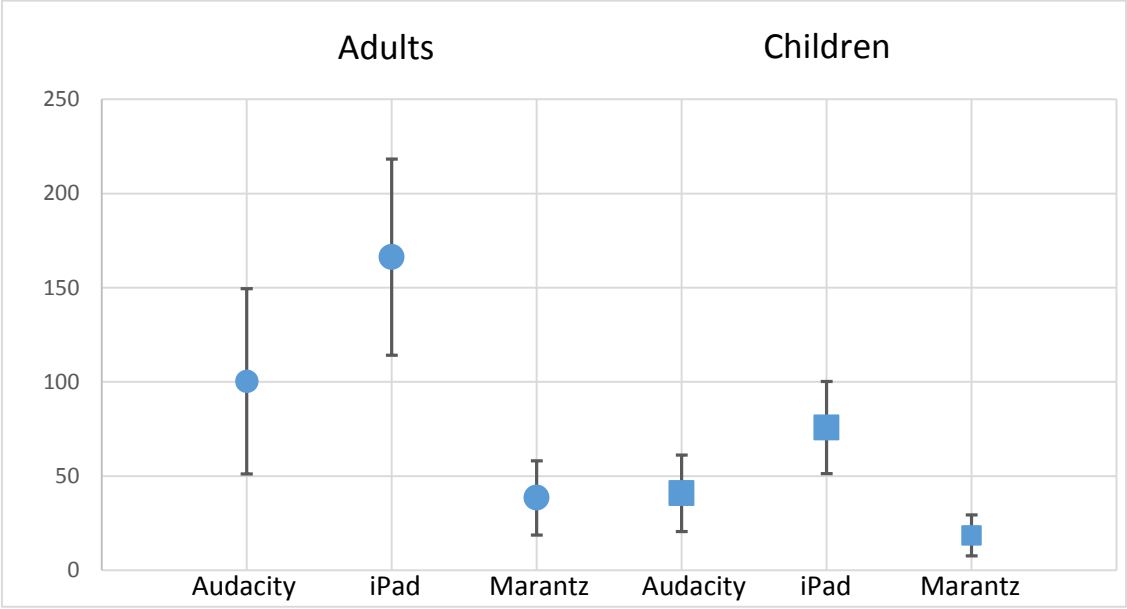
	Audacity	**iPad**	**Marantz**
Mean	100.307	166.193	38.436
Standard Deviation	58.813	62.2	23.597
95% Confidence Interval	51.13-149.48	114.18-218.2	18.7-58.17
Standard Error	49.18	52.01	19.74

Table 6: SNR for Children Recorded during the Word List Task

	Audacity	**iPad**	**Marantz**
Mean	40.857	75.765	18.535
Standard Deviation	24.272	29.118	12.918
95% Confidence Interval	20.56-61.15	51.33-100.02	7.73-29.34
Standard Error	20.3	24.44	10.81

When the word list task's recordings were separated into adults and children both samples had the highest signal-to-noise ratio from the iPad. Once again, these values were also the most variable. The Marantz again produced the lowest average SNR for both adults and children, and was again the least variable. The adult sample had higher mean SNR values on all three recording systems for this task. The values for adult and children in the word list task are plotted in Figure 2. As shown in this figure, the confidence intervals for most of the recording options overlapped for both the adult and child participants. However, recordings made with the iPad did have a better SNR than recordings made with the Marantz recorder with no overlap of the 95% confidence intervals. This finding was true for both children and adults for recordings of the word list task.

Figure 2: Means and 95% Confidence Intervals for SNR from the Word List Task



Conversational Speech Sample. Table 7 shows the overall SNRs for recordings of conversations with the children and adults. For the conversational speech sample task, the overall mean SNR for recordings from Audacity was 55.896 (*SD* = 47.063). The iPad recordings of this task had a mean of 66.331 (*SD* = 29.334). The mean for the recordings from the Marantz device was 28.265 (*SD* = 32.149). Again, the iPad recordings had the highest average SNR. However, the recordings taken by the Audacity app for PC were more variable for this event. This set of iPad recordings was the least variable of the three devices. The Marantz again had the lowest overall mean of SNR.

Table 7. Overall SNR Results for Recordings of Conversations with the Children and Adults

	Audacity	**iPad**	**Marantz**
Mean	55.896	66.331	28.265
Standard Deviation	47.063	29.334	32.149

When means and standard deviations were calculated separately for adults and children, the performance of the devices was different. The mean SNR from Audacity for adult speakers was 66.549 (*SD* = 59.51). The confidence interval for this set was 11.51-121.59 and the SEM was 55.04. The Audacity average for child speakers was 46.575 (*SD* = 34.375), resulting in a confidence interval of 17.83-75.32 and a SEM of 28.74. The iPad recordings of adult speakers for the conversational sample had a mean SNR of 62.165 (*SD* = 29.229), resulting in a confidence interval of 35.13-89 and a SEM of 27.04. The child iPad recordings had a mean SNR of 69.976 (*SD* = 30.923). This confidence interval was 44.12-95.83 and the SEM was 25.86. The Marantz device produced recordings with a mean SNR of 23.478 (*SD* = 13.387) for adults, resulting in a confidence interval of 11.1-35.86 and a SEM of 12.38. For the children, the mean from the Marantz was 35.315 (*SD* = 42.881). The confidence interval for children from the Marantz was 0.54-71.17 and the SEM was 34.77. The data for the conversation recordings is represented in Tables 8 and 9 and plotted in Figure 3.

The recordings of adult conversational speech samples were the only data set in which the iPad did not have the highest mean SNR. The mean of the Audacity recordings was highest for this group, but values were highly variable among adult participants in the sample. The data for the children was consistent with the trend in the other sets, having the highest mean SNR from the iPad. However, Marantz recordings were most variable for the children.

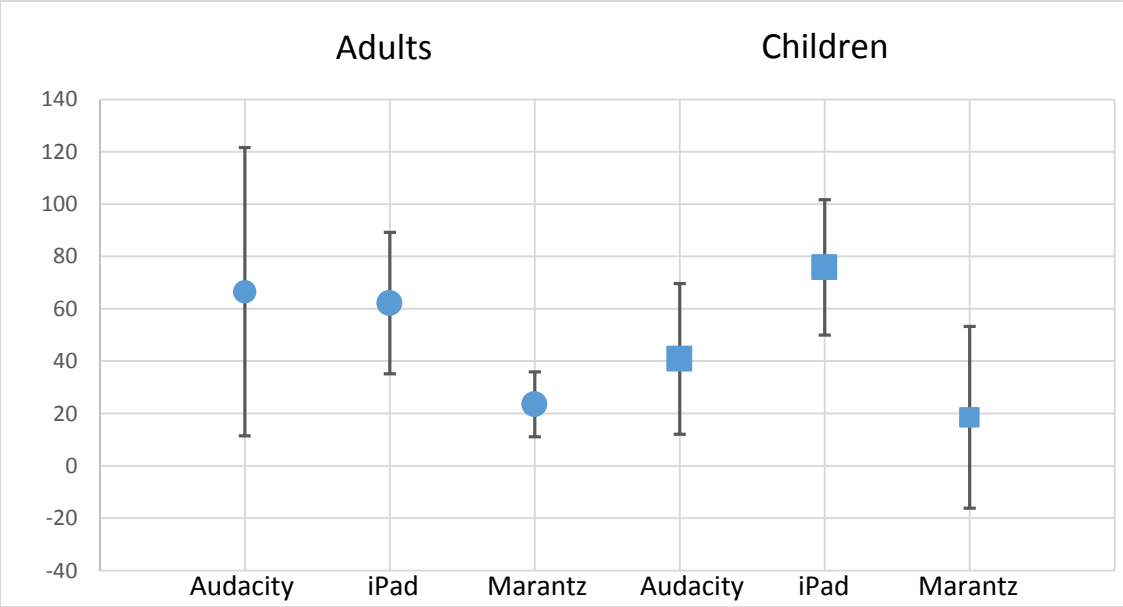
Table 8: SNR for Adults Recorded While Participating in Conversations

	Audacity	**iPad**	**Marantz**
Mean	66.549	62.165	23.478
Standard Deviation	59.509	29.229	13.387
95% Confidence Interval	11.51-121.59	35.13-89.2	11.1-35.86
Standard Error	55.04	27.04	12.38

Table 9. SNR for Children Recorded While Participating in Conversations

	Audacity	**iPad**	**Marantz**
Mean	46.575	69.976	35.315
Standard Deviation	34.375	30.923	42.881
95% Confidence Interval	17.83-75.32	44.12-95.83	0.54-71.17
Standard Error	28.74	25.86	34.77

Figure 3: Adults and Children-Conversation



Also as shown in Figure 3, the confidence intervals for all of the recording options overlapped for both the adult and child speakers. For the adult speakers the mean SNR was greatest for recordings made with Audacity on the laptop computer, but the large SEM resulted in overlap with both the iPad and the Marantz options. For children the mean SNR was greatest for iPad recordings, but again the SEM resulted in overlap with both the Audacity and Marantz options.

Clipping

Each recording was examined for the occurrence of clipping and instances were counted. Patterns were looked for to determine if clipping was more common among adults or children and males or females. It was also considered that clipping may be more likely to occur in a particular speech event. Of the 144 analyzed recordings, seven featured instances of clipping, all from children. Clipping occurred on two recordings of the Caterpillar Passage: the Audacity and iPad recordings for the same participant. The remaining five instances of clipping occurred in conversational speech samples: three on iPad recordings, one on an Audacity recording, and one on a Marantz recording. Child 001 was clipped in four of her nine recordings, child 002 in two, and child 008 in one. Two of the children were female and one was male. Neither male nor female adult participants produced signals that resulted in clipping, though the men tended to produce more intense signals, suggesting greater risk for clipping. The clipping analysis results for the children are displayed in Table 10.

Table 10. Instances of Clipping across Recordings of the Child Participants

	Audacity	iPad	Marantz
Child 001	2: Caterpillar and Conversation	2: Caterpillar and Conversation	0
Child 002	0	1: Conversation	1: Conversation
Child 003	0	0	0
Child 004	0	0	0
Child 005	0	0	0
Child 006	0	0	0
Child 007	0	0	0
Child 008	0	1: Conversation	0

Discussion

Recording Technology

One purpose of the present study was to determine what combination of microphones and recording devices would provide the clearest speech sample based on a signal-to-noise ratio and absence of peak clipping or signal distortion. Though there were differences in signal-to-noise ratio for each speaking task across the three recording systems, values varied greatly and in many cases ranges of values overlapped across devices. There was no one device and microphone pairing that was clearly superior to the others, suggesting that they are capable of recording signals of similar quality. The age of the participants and the speaking tasks also were factors a speech-language pathologist would need to consider. For the reading passage task, the SNR for the iPad was clearly higher than the SNR for the Marantz device in the recordings from the adult speakers. For the child speakers, the SNR results across all three recording options overlapped to

a considerable degree and no one recording option stood out as the best one. One possible explanation for these differences is that some of the child speakers were not proficient readers and completed the reading passage task by repeating the sentences rather than by reading the sentences. This might have resulted in a speech sample with different characteristics than an actual passage reading. For the word list task the iPad stood out for both the child and adult speakers as having a higher SNR than the Marantz recorder. As with the reading passage task, SNRs for the Audacity and iPad recording tended to overlap. None of the devices stood out as having better SNRs for the recorded conversations. One of the interesting findings was that peak clipping did not occur in any of the adult speech samples and only in seven instances among the children.

The findings of this study suggested that readily available devices, when paired with an appropriate microphone, yielded speech recordings that were as good or better than the recordings made with the dedicated recording device. Many speech-language pathologists and researchers in the field have personal computers or iPads, or have those devices available to them within their work or educational institutions. However, not all of these individuals are likely to have access to a dedicated recording device. Apps paired with the PC and iPad were free and easy to use and appeared to provide recordings that were adequate for the phonetic transcription and some other speech sample analyses. Any recommendations based on the findings of this study are tentative. One unexpected finding was the high degree of variability across participants that yielded large confidence intervals. The sources of this variability warrant further investigation. The Marantz recorder also proved to be more difficult to operate consistently. The researcher inadvertently recorded 5 of the 48 samples using the MPEG3 format instead of the .wav format selected for the study. These 5 instances all occurred with the Marantz recorder.

Additionally, newer models of all three pieces of equipment: the HP computer, iPad, and Marantz are available. Recordings taken on newest models could potentially produce signals clearer than those acquired in this study.

Group Differences

Another purpose of the present study was to determine if the quality of recordings for the different devices depended on the participant differences such as age and gender identity. Due to limitations in the adult sample, the researcher did not pursue the gender comparisons. Sample size was small and because of a transmission issue, one of the male's recordings were missing from analyses, resulting in an unequal number of males and females. However, the researcher did make comparisons across the adult and child groups. Peak clipping was one of the measures employed in this study and the results revealed peak clipping only in the child samples. Peak clipping occurs when the amplitude of a signal exceeds the level that recording device and microphone can handle. Before making recordings, the researcher adjusted the recording levels for each device to achieve optimal levels. Apparently, these adjustments were less accurate for child speakers. Another measure used in this study was signal-to-noise ratio (SNR). Some differences occurred across the adult and child speakers. In the reading passage task the SNRs for adult and child speakers overlapped but the results for the child speakers were more variable, based on standard errors of measurement and confidence intervals. For the word list task, the iPad recordings of adult speakers yielded higher SNRs than the iPad recordings of child speakers. For this task, the adult speakers had more variable results than the child speakers. Finally, the conversational speech task yielded no clear differences in SNR between the adult and child speakers for any of the recording devices.

Environment

Much of the variability in the results of SNR analysis can be attributed to the difficulty of controlling the recording environment. Even after instruction, child participants had a difficult time sitting still and maintaining a constant distance from the microphones. Each session was not recorded in the same therapy room of the speech and hearing clinic where the research took place. Room acoustics could contribute to the signal and noise values of recordings. Speakers produced sounds of greater intensity during the speech tasks than during the meter check, making clipping possible. High frequency sounds, like fricatives, were produced with greater intensity within the speech tasks. Also, natural rises in intensity occurred during the conversational samples as many speakers showed enthusiasm in their responses. Clipping resulted in a displayed signal value of 1 for the recordings in which it occurred, resulting in a higher signal-to-noise ratio and raising the mean for the sample. Without clipping, average values would have been slightly different, which may have altered comparison within the events that featured clipping. These variables are similar to those that would be encountered when recording during an evaluation, therapy session, or research study.

Future Research

Further investigation into optimal microphone pairings could provide higher quality for all recording technologies. If head mounted microphones were used according to National Center for Voice and Speech's recommendations, and recordings of the same speech task on different devices were taken individually, results may have been more consistent, providing a clearer comparison. In this study, analyses focused on basic acoustic features. This may have contributed to the lack of a clearly superior recording apparatus, as the study by Vogel and Maruff (2008) suggested. Future research could investigate why the Marantz SNR values were

more consistent during the reading passage and word list tasks, and if this consistency were necessary for more sophisticated acoustic analysis.

Conclusions

Transcription and reevaluation of speech samples is important to students and professionals in the field of communication disorders for research, instruction, and evaluation. Recordings of speech samples allow speech-language pathologists to focus on eliciting the intended sample and maintaining the environment and interaction for the client. Professional speech recorders intended for research are expensive and have few uses. More common technologies, such as personal computers and iPads, are capable of recording speech samples and can be used for a variety of purposes in both clinical and educational settings.

This study analyzed recordings taken by three different devices: a Marantz PMD 660 dedicated recorder, an HP personal computer with Audacity application, and a generation 4 iPad with Voice Record Pro app paired with three omnidirectional, condenser lavalier microphones. Recordings featured three speech tasks: a standard passage, word list, and conversational speech sample, produced by adult and child participants. Following analysis for signal-to-noise ratio and clipping using Praat software, it was concluded that none of the recording systems were clearly superior to the others. Though data sets were varied, each technological device was capable of producing high quality signals, indicating that readily available devices can be used for these types of analyses.

Speech-language pathologists use laptop computers and tablets for a variety of functions, including therapy activities. This study indicated that these devices are suitable for basic acoustic analyses used to examine clients' speech. Therefore, their use could be extended to include recordings of speech samples, resulting in lower costs and greater efficiency for professionals.

References

American Speech-Language Hearing Association (2016). Scope of practice in speech-language pathology. *American Speech-Language Hearing Association.*

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

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

Deliyski, D., Evans, M., & Shaw, H. (2004). Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice, 19*(2), 176-186.

Howard, D.M., & Murphy, D.T. (2007). *Voice science acoustics and recordings*. San Diego, CA: Plural Publishing.

Huckvale, M. (2009). *Equipment for audio recording of speech.* Retrieved from http://www.phon.ucl.ac.uk/resource/audio/recording.html

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

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

Small, L. H. (2016). *Fundamental of phonetics: A practical guide for students* (3rd ed.). Boston, MA: Pearson.

Svec, J., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology, 19*, 356-368.

van Lieshout, P. (2003). Praat short tutorial: A basic introduction. Retrieved from https:// web.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf

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

Appendix A

Informed Consent for Adult Participants

Project Title: A Comparison of Technologies for Recording Speech

Name of Investigator(s): Jacquelyn Knustrom

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 provide evidence as to whether or not non-professional recording devices can produce recordings similar in quality to professional recording devices and therefore be used as a cheaper, more readily available apparatus for obtaining speech samples for instructional purposes. Acquiring quality speech samples plays a vital role in transcription for Speech-Language Pathologists to accurately assess errors in production of speech sounds.

Explanation of Procedures: You will be asked to respond to several background questions before participating. This will include identifying your age, gender, hearing status, and history of speech disorders. You will then read a word list, read a short passage, and engage in a conversational narrative with the investigator. These speech samples will be recorded on three separate devices so the quality of recordings can be compared. Your recording will be labeled by your age and gender so the investigator can determine if those factors affect the quality of recording. Your recording will be completed in a single session. Recordings will be stored securely in the Roy Eblen Speech and Hearing Clinic following the conclusion of the study for possible instructional use. Age and gender will be the only identifying information saved with recordings.

Discomfort and Risks: Risks to participation are similar to those experienced in day-to-day life.

Benefits and Compensation: You will receive no direct benefits or compensation for your participation.

Confidentiality: Information obtained during this study which could identify you will be kept confidential. The summarized findings with no identifying information may be published in an academic journal or presented at a scholarly conference. Recordings will be stored in a secure area of the Roy Eblen Speech and Hearing Clinic on a password protected USB drive. In the future recordings will be used only for academic research or instruction. Your name and identification number will not be available with the recording following the conclusion of this study.

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.

Questions: If you have questions about the study or desire information in the future regarding your participation or the study generally, you can contact Jacquelyn Knustrom at 319-316-2034. 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)

_____ _____

Appendix B

Informed Consent for Child Participants

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

Nature and Purpose: The purpose of this research is to provide evidence as to whether or not non-professional recording devices can produce recordings similar in quality to professional recording devices and therefore be used as a cheaper, more readily available apparatus for obtaining speech samples for instructional purposes. Acquiring quality speech samples plays a vital role in transcription for Speech-Language Pathologists to accurately assess errors in production of speech sounds.

Explanation of Procedures: You will be asked to respond to several background questions on behalf of your child before he or she participates. This will include identifying his or her age, gender, hearing status, and history of speech disorders. Your child will then read or repeat a word list, read or repeat a short passage, and engage in a conversational narrative with the investigator. These speech samples will be recorded on three separate devices so the quality of recordings can be compared. Your child's recording will be labeled by his or her age and gender so the investigator can determine if those factors affect the quality of recording. Your child's recording will be completed in a single session. Recordings will be stored securely in the Roy Eblen Speech and Hearing Clinic following the conclusion of the study for possible instructional use or further research. Age and gender will be the only identifying information saved with recordings.

Discomfort and Risks: Risks to participation are similar to those experienced in day-to-day life

Benefits: Your child will receive no direct benefits or compensation for his or her participation.

Confidentiality: Information obtained during this study which could identify your child will be kept strictly confidential. The summarized findings with no identifying information may be published in an academic journal or presented at a scholarly conference. Recordings will be stored in a secure area of the Roy Eblen Speech and Hearing Clinic on a password protected USB drive. In the future recordings will be used only for academic research or instruction. Your child's name and identification number will not be available with the recording following the conclusion of this study.

Right to Refuse or Withdraw: Your child's participation is completely voluntary. He or she is free to withdraw from participation at any time or to choose not to participate at all, and by doing so, your child will not be penalized.

Questions: If you have questions about the study or desire information in the future regarding your child's participation or the study generally, you can contact Jacquelyn Knustrom at 319-

316-2034. You can also contact the office of the Human Participants Coordinator, 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 child's participation in this project as stated above and the possible risks arising from it. I hereby agree to allow my son/daughter to participate in this project. I have received a copy of this form.**

_____ _____
(Signature of parent/legal guardian) (Date)

_____
(Printed name of parent/legal guardian)

_____
(Printed name of child participant)

_____ _____
(Signature of investigator) (Date)

_____ _____
(Signature of instructor/advisor) (Date)

Appendix C

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 D

Caterpillar Passage

Do you like amusement parks? Well, I sure do. To amuse myself, I went twice last spring. My most MEMORABLE moment was riding on the Caterpillar, which is a gigantic roller coaster high above the ground. When I saw how high the Caterpillar rose into the bright blue sky I knew it was for me. After waiting in line for thirty minutes, I made it to the front where the man measured my height to see if I was tall enough. I gave the man my coins, asked for change, and jumped on the cart. Tick, tick, tick, the Caterpillar climbed slowly up the tracks. It went SO high I could see the parking lot. Boy was I SCARED! I thought to myself, "There's no turning back now." People were so scared they screamed as we swiftly zoomed fast, fast, and faster along the tracks. As quickly as it started, the Caterpillar came to a stop. Unfortunately, it was time to pack the car and drive home. That night I dreamt of the wild ride on the Caterpillar. Taking a trip to the amusement park and riding on the Caterpillar was my MOST memorable moment ever!

Appendix E

Word List

choice
swing
those
eat
witch
kid
crab
thunder
hand
teacher
family
globe
fishing
home
reward
star
yellow
zipper
spider
cow
mouse
oatmeal
taught
move
cat
pajamas
sled
leaf
present
wife
toy
shoe
vacuum
mother
pencil
toe
window
barn
blue
eye
nose
breathe
thank
goat
baby
crackers
jumping
eight
leash
smoke
off
clown
fish
itch
tiger
sky
train
bet
five
rope
weigh
beige
look
muscle
alien
behind
flag
twins
mean
up
chair
about
hour
mask
plant
teeth
oil
ask
orange
tame
quick
tree
television
green
umbrella
carrot
on
duck
bathtub
end
oops
telephone
buy
snake