

AT&T Final Report

Linda Kozma-Spytek^{}, Paula Tucker^{*}, Mary Garvert⁺, Christian Vogler^{*}*

^{}Gallaudet University Technology Access Program*

⁺Gallaudet University Hearing, Speech and Language Sciences

{linda.kozma-spytek,mary.garvert,paula.tucker,christian.vogler}@gallaudet.edu

February 22, 2016

Table of Contents

Executive Summary	1
Introduction	2
Experimental Procedure	3
Participants	3
Materials	4
Method	7
Results	10
Speech Understanding	10
Hearing Participants	10
Participants with Hearing Loss	11
Mental Effort (SMEQ)	13
Hearing Participants	13
Participants with Hearing Loss	14
Mean Opinion Score (MOS)	16
Hearing Participants	16
Participants with Hearing Loss	17
Purchase Intent	19
Hearing Participants	19
Participants with Hearing Loss	19
Conclusions	20
Acknowledgments	22
Appendix A	23
Sentence Set 1	23
Sentence Set 2	23
Sentence Set 3	24
Sentence Set 4	24
Sentence Set 5	25
Sentence Set 6	26
Appendix B	27
Source Stimuli	27
Test Conditions	27
Preprocessing Steps (all)	27
Preprocessing Step (for each condition/sentence set combination)	27
Processing Steps- AMR-WB	27
Processing Steps- AMR-NB	27
Total number of processed files to be delivered by ATT	28
Postprocessing steps (all)	28
Appendix C	29
Speech Understanding	29
Speech Quality	29
Subjective Mental Effort	29
Purchase Intent	30
Appendix D	31

Data Table for Hearing Participants..... 31
Data Table for Participants with Hearing Loss..... 32
ANOVA: Two Within-Subject Factors..... 33
McNemar’s Test: Purchase Intent..... 34

Executive Summary

The purpose of this study was to broaden our understanding of the effects various technical parameters can have on audio-only voice communications in cellular and VoIP telephony environments by individuals with hearing loss. The impact of audio-bandwidth and packet loss on speech understanding, mental effort, sound quality and purchase intent was examined using a double-blind, with-in subjects/repeated measures experimental design. Thirty-six individuals with hearing loss and twelve hearing individuals participated in one-hour test sessions in which they listened to sets of twenty sentences each under conditions of narrowband and wideband audio coding with 0%, 3% and 20% levels of packet loss. The results suggest that wideband audio improves the overall quality of experience for both hearing individuals and individuals with hearing loss compared to narrowband audio, especially when network impairments due to packet loss are low. However, at the highest level of network impairment, individuals with hearing loss experienced appreciably greater declines across all dependent measures than their hearing counterparts suggesting a limit to any wideband audio advantage for this group.

Introduction

Under the auspices of the Rehabilitation Engineering Research Center on Telecommunications Access, the Technology Access Program (TAP) at Gallaudet University has had an ongoing program of research on voice telecommunications accessibility for individuals with hearing loss over the past five years. The goal of this program is to better understand the technical parameters of cellular and VoIP telephony environments that lead to effective audio-only and audio-visual voice communications by individuals with hearing loss. Previously, five with-in subject experiments, of approximately 120 individuals with hearing loss, have been completed. These experiments have investigated the effects of presentation mode (audio-only and the addition of a video stream), video quality (video frame rate and audio-video synchrony), audio quality (codec audio bandwidth and data rate) and the receive environment (quiet and the addition of noise), under both simulated and actual wireless device use.

The goal of the research described in this report was to investigate the impact of packet loss on audio quality, as perceived by people with hearing loss, for both narrowband and wideband audio (also called HD Voice). In particular, the current experiment expands the previous line of research to include a new audio quality condition involving the network impairment of packet loss, and at the same time replicates a previously examined audio quality condition, codec audio bandwidth. This report describes the study participants and the materials and methods of the experimental protocol. The results are summarized using both descriptive and inferential statistics. The conclusion provides a discussion of the results from this experiment and, next steps are proposed for future research directions.

While the current experiment utilizes the same basic experimental protocol from the previous experiments, several aspects have been adjusted to either strengthen or extend the protocol. To strengthen the protocol, a double blind procedure was used. Double blind procedures serve to protect against certain types of experimental bias. Additionally, more difficult sentence material was used to reduce the likelihood of a ceiling effect in speech understanding, and the number of speakers represented in the stimuli was increased. The protocol was extended to include a group of individuals without hearing loss. This provides a means of direct comparison of results between individuals with hearing loss and hearing individuals, whose quality of experience is most often considered by industry.

Experimental Procedure

Participants

Twelve hearing individuals and thirty-six individuals with hearing loss participated in the study. All participants were fluent English speaking adults, 18 years of age or older; hearing participants were all younger than 50 years old. All participants passed a hearing screening for audibility of the higher frequencies (4 kHz and 5 kHz) contained within wideband telephony; for participants with hearing loss the screening was completed while using their hearing devices. Participants with hearing loss were also required to be daily hearing aid or cochlear implant users and regular users of the voice telephone. Participants were solicited through the Hearing Loss Association of America and Gallaudet University.

Of the 36 individuals with hearing loss, 25 were women and 11 were men, with an average age of 51 years (min. age: 22 years; max. age: 79 years). All participants had at least two years self-reported hearing device use. Twelve individuals used their cochlear

implants during testing, while the other 24 used their hearing aids. Self-reported hearing loss ranged from mild to profound across both ears. In the test ear, most hearing aid users reported moderately-severe or severe hearing loss, while all cochlear implant users reported profound hearing loss.

Materials

Stimuli for this experiment were drawn from the IEEE Harvard sentence lists. The Harvard sentences are a collection of 72 lists of 10 sentences/phrases that are phonetically balanced, using specific phonemes at the same frequency they appear in English. Because these lists date back to the 1940s and language use patterns have shifted since then, lists containing words that might have been offensive or unfamiliar to participants today were screened out. In the end, sentence lists 11, 14, 15, 16, 17, 19, 21, 22, 23, 25, 26 and 29 were used to prepare the stimuli. Pairs of lists were combined to create six sets of 20 sentences each with, on average, 157 words per set. The exact sentences for each set are given in Appendix A.

Recordings of all sentences were obtained from Harry Levitt of Sense Synergy and included four speakers (2 male and 2 female) per sentence. Within a set of 20 sentences in a single test condition, five sentences were spoken by each of the four speakers, and each participant received a different mix of speakers across all conditions. Furthermore, the presentation order of sentences in each set was randomized across test conditions and participants. Additionally, other IEEE Harvard sentences were used to train participants on the procedure and to establish an individual's most comfortable listening level (MCL) for telephone listening.

Two baseline conditions and four conditions of reduced audio quality due to packet loss were prepared (Table 1). The baseline conditions included two audio-bandwidth encoding strategies: 1) Adaptive Multi-Rate-Narrowband (AMR-NB) @ 5.90 kbps and 2) Adaptive Multi-Rate-Wideband (AMR-WB) @ 12.65 kbps. Typical data rates used in mobile cellular networks determined the bit rates, which were selected for each audio codec in consultation with AT&T. Compared to the narrowband audio codec, the wideband codec extends both the upper and lower limits of the frequency bandwidth by doubling the upper limit from 3400 Hz to 7000 Hz and extending the lower frequency limit down from 200 Hz to 50 Hz. However, practically, the lower frequency limit will be circumscribed by the receiver characteristics of the handset, and the upper frequency limit will be constrained by the codec's in-use data rate.

The experimental stimuli of reduced audio quality were developed under two levels of bursty packet loss, 3% and 20%, for each baseline condition. The worst-case quality levels in a managed mobile cellular network and over unmanaged Wi-Fi Internet calling, respectively, determined the percentages of packet loss selected. These test conditions provided upper and lower boundaries for the speech quality levels due to the effects of packet loss likely to be experienced in mobile calling. AT&T Labs carried out all signal processing for speech coding and injection of packet loss.

Table 1: Experimental Conditions of Audio Codec and Packet Loss

Condition	Audio Codec			Packet Loss		
#	Compression Format	Audio Bandwidth	Bit Rate	Model	Percent	Gamma ⁴
1	AMR ¹	NB ²	5.90	Gilbert-Elliott	0%	---
2	AMR	NB	5.90	Gilbert-Elliott	3%	0.8

3	AMR	NB	5.90	Gilbert-Elliott	20%	0.8
4	AMR	WB ³	12.65	Gilbert-Elliott	0%	---
5	AMR	WB	12.65	Gilbert-Elliott	3%	0.8
6	AMR	WB	12.65	Gilbert-Elliott	20%	0.8

¹Adaptive Multi-Rate (AMR) – DTX and therefore, VAD and CNG were disabled, PLC was provisioned on

²narrowband (NB) – AMR-NB encodes 200 – 3400 Hz

³wideband (WB) – AMR-WB encodes 50 – 7000 Hz

⁴Gamma is the variance in error burst length. The distribution of state durations/lengths are based on the theoretical probability Gamma distribution; such that as the variance in burst length increases the burstiness of packet loss increases and conversely, as the variance in burst length decreases the burstiness of packet loss decreases.

To prepare the stimuli for processing, the silences at the beginning and end of each sentence were deleted, and the sentences within a given set were concatenated together. Each concatenated sentence set was processed using the two encoding strategies. Silence suppression (DTX) was hardcoded to be off in both the narrowband and wideband codec implementations used to process all stimuli, which means comfort noise generation and voice activity detection were also disabled. Since the silence preceding and following each sentence was removed before the sentences in each set were joined together, it is unlikely there is any impact of having DTX off. Packet loss concealment (PLC) was provisioned on, and the same technique, a form of waveform substitution, was used for both codec implementations. Basically, with this technique, the last received packet is repeated in place of each lost packet until another packet is received. When more than one packet is lost in a row, as is the case with bursty packet loss, the signal level of each substituted packet is reduced. This reduction continues progressively for each lost packet until a level approximating that of comfort noise is reached or a new packet is received.

Packet loss was introduced using the Gilbert-Elliott model with Gamma set at 0.8. The Gilbert-Elliott model utilizes a two-state Markov model approach and is widely

used to generate impairments that simulate transmission failures in real-time services over telecommunications networks. Within this model, lower levels of Gamma produce more random packet loss distributions, while higher levels of Gamma produce more bursty distributions of packet loss. Packet loss in both mobile and VoIP networks has been characterized as bursty, rather than random. Therefore, a higher level of Gamma was selected in order to simulate the bursty nature of packet loss in these telephony environments. Following processing, the sentences in each set were separated and approximately 100 ms of silence was added to the beginning and end of each sentence. The step-by-step procedure for stimuli preparation can be found in Appendix B.

All sentence sets were processed for each of the six test conditions, and the sentence set used for each condition was counterbalanced across subjects. This was done to guard against the effects of possible differences in intelligibility, either inherent or as a result of the temporal distribution of lost packets, among sentence sets.

Method

Participants' preferred ear and self-selected speech MCL for telephone listening were used in all test conditions. An iPhone 4S was used for presentation of stimuli. A custom app developed by TAP was used to control presentation of all stimuli in the correct order, and to control the phone settings. No cellular or WiFi network connections were active on the phone during testing. The phone was placed in a normal use position at a participant's ear for hearing individuals and at the microphone of a participant's hearing device for individuals with hearing loss. An adjustable stand was used to position and hold the handset. A Velcro headband was loosely placed around the participant's head and the phone to assist the listener in maintaining the relative positioning of their ear or hearing

device's microphone and the phone's speaker for best-case acoustic coupling (Figure 1). A Bluetooth keyboard paired with the phone was used by the testers to interact with the phone; no on-phone buttons were used. Prior to the start of testing, all participants received training on the entire procedure, with instructions provided both verbally and in writing.

The speech MCL was established at the beginning of testing and locked on the phone for the remainder of testing. While participants held the phone to their ear or hearing device's microphone, the volume control (VC) setting of the phone was set at its mid point. Each participant then listened to the telephone speech indicating either verbally or with hand movements whether they wanted the tester to increase or decrease the VC via the Bluetooth keyboard so that the speech level was comfortably loud. The VC setting was adjusted up and down several times to converge on a consistent MCL judgment. The phone was then placed in the stand. The VC setting for the MCL judgment was confirmed, and the setting was locked.

Figure 1: Experimental setup



For each study participant, speech understanding was tested using one set of IEEE sentences for each of the six audio quality conditions. Participants listened to and then repeated each sentence that they heard, and TAP staff scored their responses for the number of words correctly repeated in each sentence. Following presentation of all 20 sentences for a given condition, the Subjective Mental Effort Questionnaire (SMEQ) and the Mean Opinion Score (MOS) were administered. The order of administration of the SMEQ and MOS was counterbalanced across subjects. The SMEQ provides a post-task rating of the mental effort an individual expends in completing a task. It consists of a single scale with values from 0 to 150 and nine labels from “Not at all hard to do” to “Tremendously hard to do.” Participants moved a slider with their finger to the point in the scale that represented their judgment of task difficulty. The slider widget calculated and provided the scale value selected by the participant. The MOS is an absolute category rating for speech quality on a 5-point scale from excellent/5 to bad/1. Participants selected the category which best

represented the overall quality of speech they experienced when listening to the sentences for a given condition. Then, each participant answered a yes/no question regarding whether they would purchase and use a cell phone with the level of sound quality represented in the sentence set just completed. The dependent measures as implemented in the experiment can be found in Appendix C. Lastly, audibility of third octave band noises centered at 150, 250, 4k, 5k Hz was tested to reconfirm the results of the audibility screening individuals were required to pass in order to participate in the study.

Each administration of one condition took approximately six to seven minutes. Presentation of conditions was counterbalanced across subjects. To guard against bias, a double blind procedure was used in which neither the TAP staff administering the experiment nor the participants were aware of which conditions were being evaluated for any given sentence set. Each testing session lasted approximately one hour.

Results

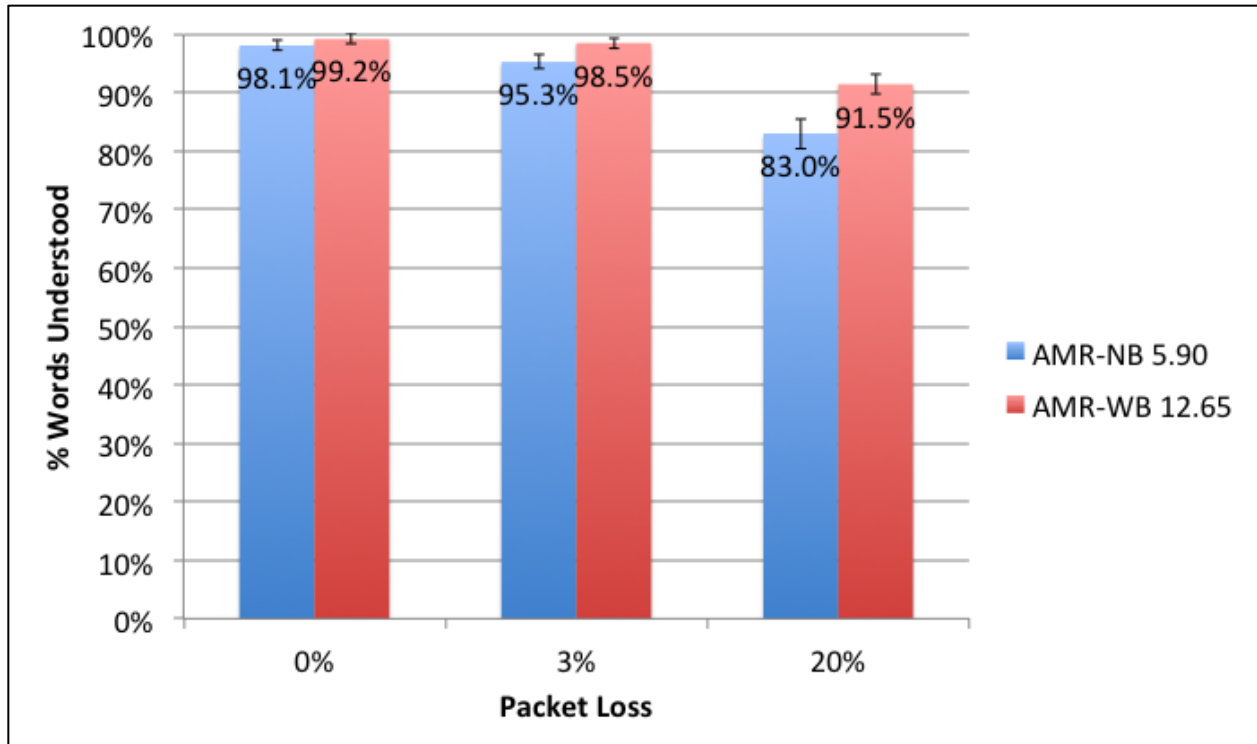
Speech Understanding

Hearing Participants

Hearing participants had near perfect levels of speech understanding, 98% and 99% words correct, for both narrowband (NB) and wideband (WB) telephone speech, respectively, at 0% packet loss. A slight reduction in speech understanding (~3 %age points) occurred for NB speech when packet loss was increased to 3%, with no like reduction in speech understanding for WB speech at 3% packet loss. When packet loss was increased to 20%, speech understanding was further reduced by ~12 %age points for NB speech and reduced by ~9 %age points for WB speech (Figure 2).

For the hearing participants, a repeated measures, two-way ANOVA for words correct showed significant ($\alpha=0.05$) main effects of the factors audio bandwidth ($F_{(1,11)} = 16.0, p<0.002$) and packet loss ($F_{(2,22)} = 31.9, p<0.000$) and a significant interaction between the two factors ($F_{(2,22)} = 5.82, p<0.009$).

Figure 2: Speech Understanding for Hearing Participants (n=12)



Participants with Hearing Loss

Participants with hearing loss had limited levels of speech understanding. Within these limits, WB speech understanding was higher than NB speech understanding across all levels of packet loss. At 0% and 3% packet loss, the difference in speech understanding between WB and NB speech was ~4 %age points. This difference approximately doubled at 20% packet loss, although average speech understanding was low, ~43% and 51% respectively, for both NB and WB speech (Figure 3). Overall, participants with hearing loss

had poorer speech understanding than hearing participants, regardless of audio codec bandwidth and degree of packet loss (Figure 4).

For participants with hearing loss, a repeated measures, two-way ANOVA for words correct showed significant ($\alpha=0.05$) main effects of the factors audio bandwidth ($F_{(1,35)} = 16.8, p<0.000$) and packet loss ($F_{(2,70)} = 278, p<0.000$), but no significant interaction between the two factors ($F_{(2,70)} = 1.44, p<0.243$).

Figure 3: Speech Understanding for Participants with Hearing Loss (n=36)

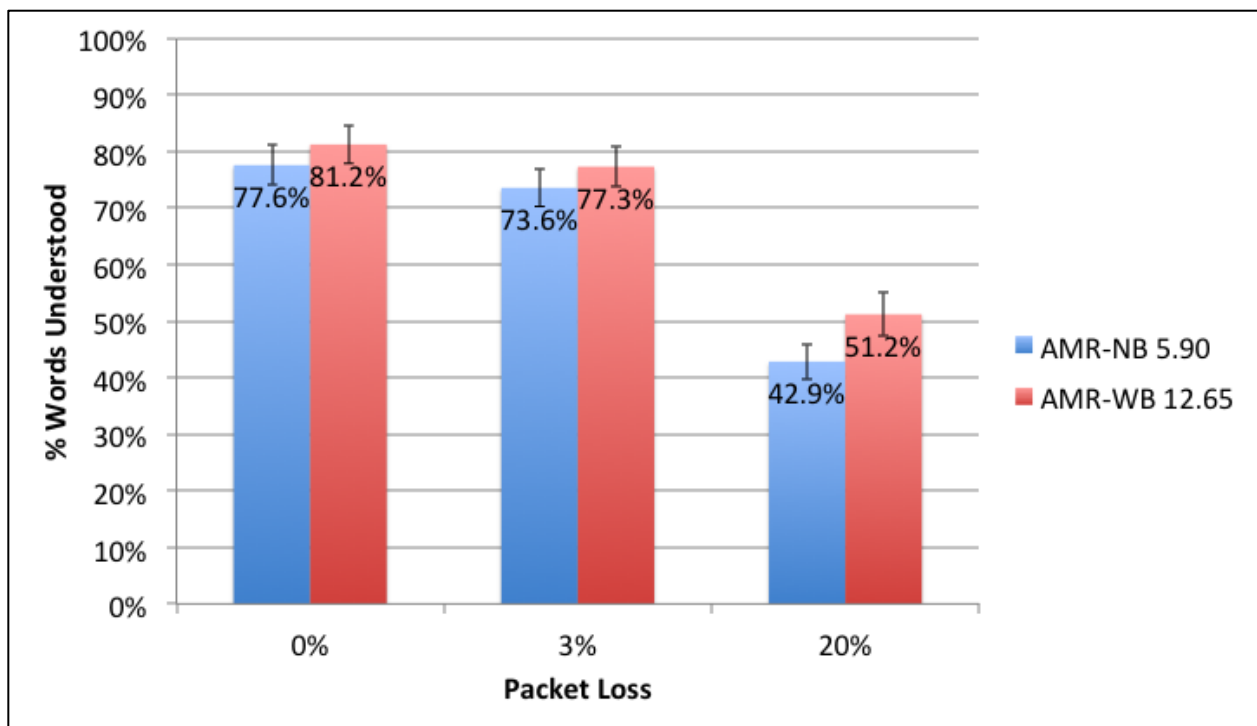
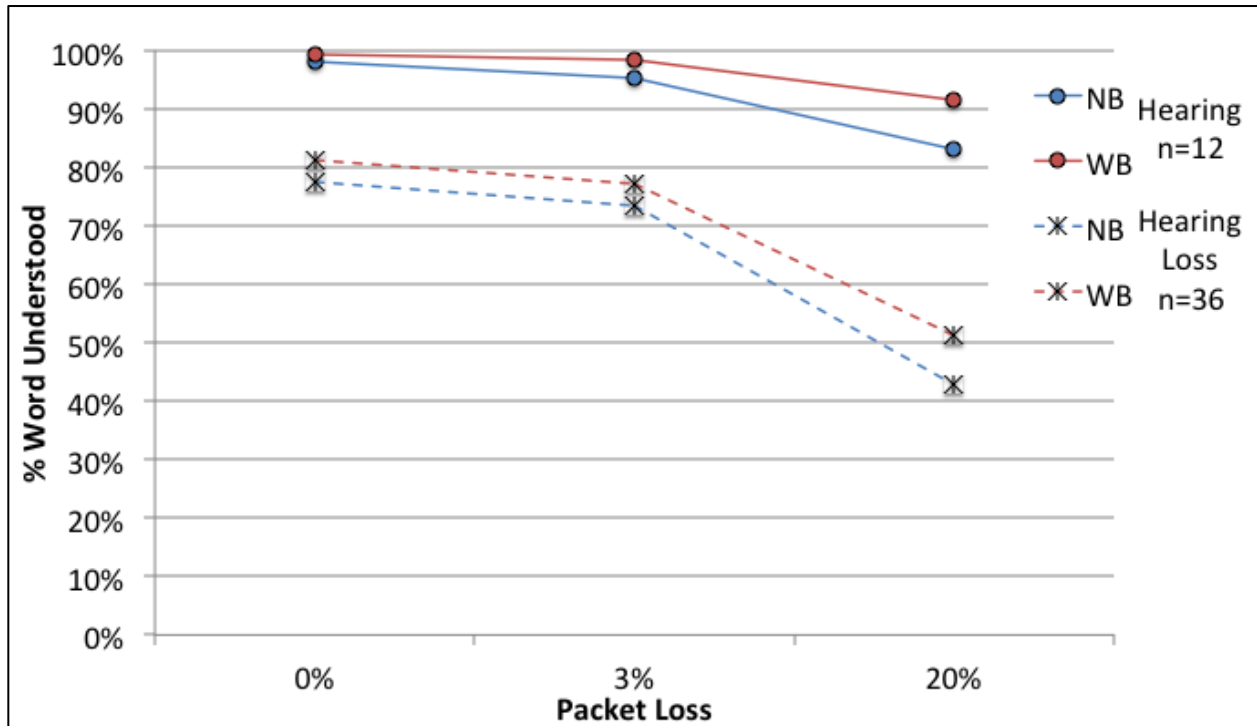


Figure 4: Speech Understanding for Hearing Participants and Participants with Hearing Loss



Mental Effort (SMEQ)

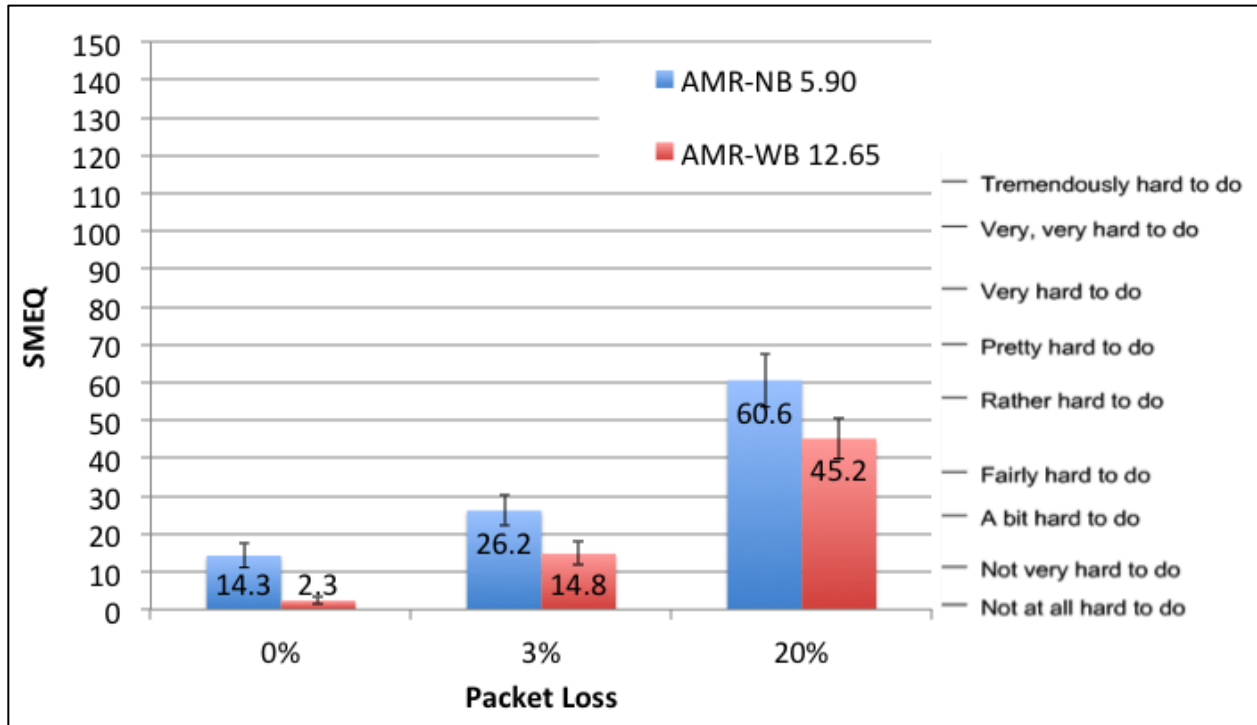
Hearing Participants

Hearing participants expended higher levels of mental effort to understand NB speech compared to WB speech across all levels of packet loss. Expenditures of mental effort increased as packet loss increased for both NB and WB speech. At the highest level of packet loss, hearing participants reported that the speech understanding task was fairly hard to do when listening to WB speech and rather hard to do when listening to NB speech. Without any packet loss, these same participants reported that the speech understanding task was not at all hard to do for WB speech and not very hard to do for NB speech (Figure 5).

For the hearing participants, a repeated measures, two-way ANOVA for SMEQ ratings showed significant ($\alpha=0.05$) main effects of the factors audio bandwidth ($F_{(1,11)} =$

14.0, $p < 0.003$) and packet loss ($F_{(2,22)} = 47.3$, $p < 0.000$), but no significant interaction between the two factors ($F_{(2,22)} = 0.329$, $p < 0.723$).

Figure 5: Subjective Mental Effort for Hearing Participants



Participants with Hearing Loss

Participants with hearing loss had to expend substantial mental effort for all tasks. Expenditures of mental effort increased as packet loss increased from 0% to 3% for both NB and WB speech. At the highest level of packet loss, participants with hearing loss reported that the speech understanding task was very hard to do for both NB and WB speech. Without any packet loss, these same participants reported that the speech understanding task was fairly hard to do for WB speech and rather hard to do for NB speech (Figure 6). Overall, participants with hearing loss had higher expenditures of mental effort for the speech understanding task than hearing participants, regardless of audio codec bandwidth and degree of packet loss (Figure 7).

For participants with hearing loss, a repeated measures, two-way ANOVA for SMEQ ratings showed significant ($\alpha=0.05$) main effects of the factors audio bandwidth ($F_{(1,35)} = 4.44, p<0.042$) and packet loss ($F_{(2,70)} = 54.1, p<0.000$), but no significant interaction between the two factors ($F_{(2,70)} = 1.75, p<0.181$).

Figure 6: Subjective Mental Effort for Participants with Hearing Loss

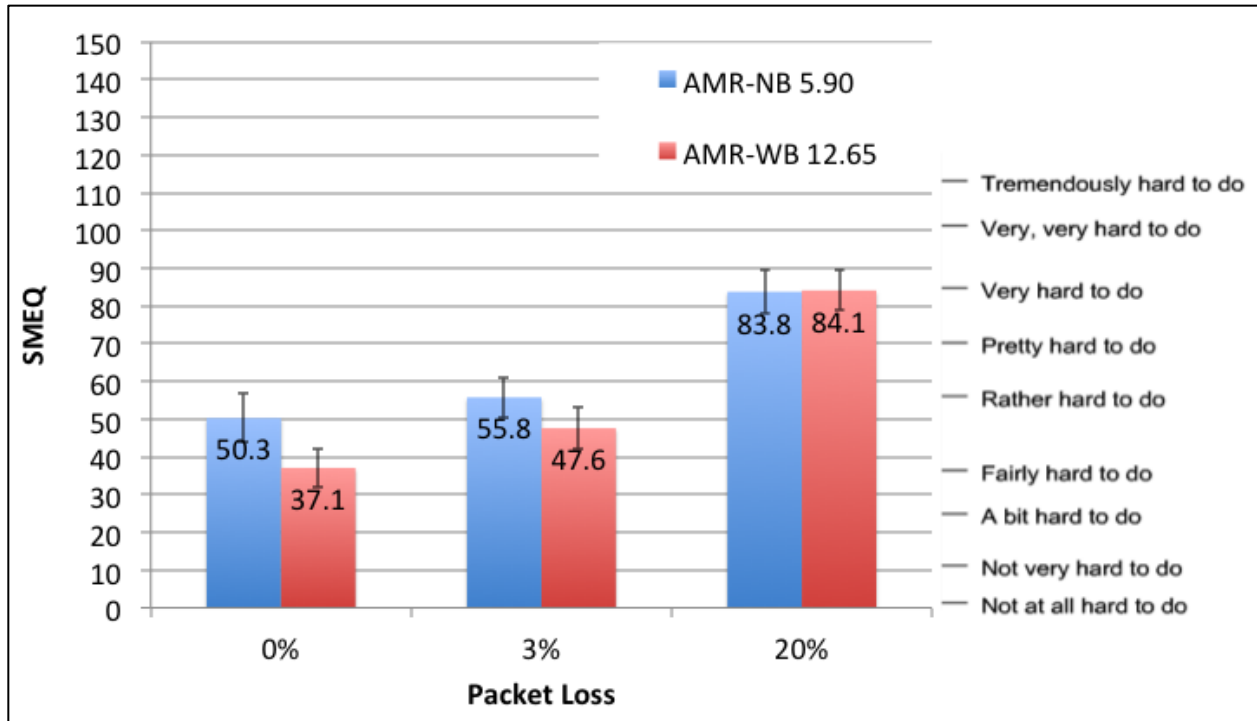
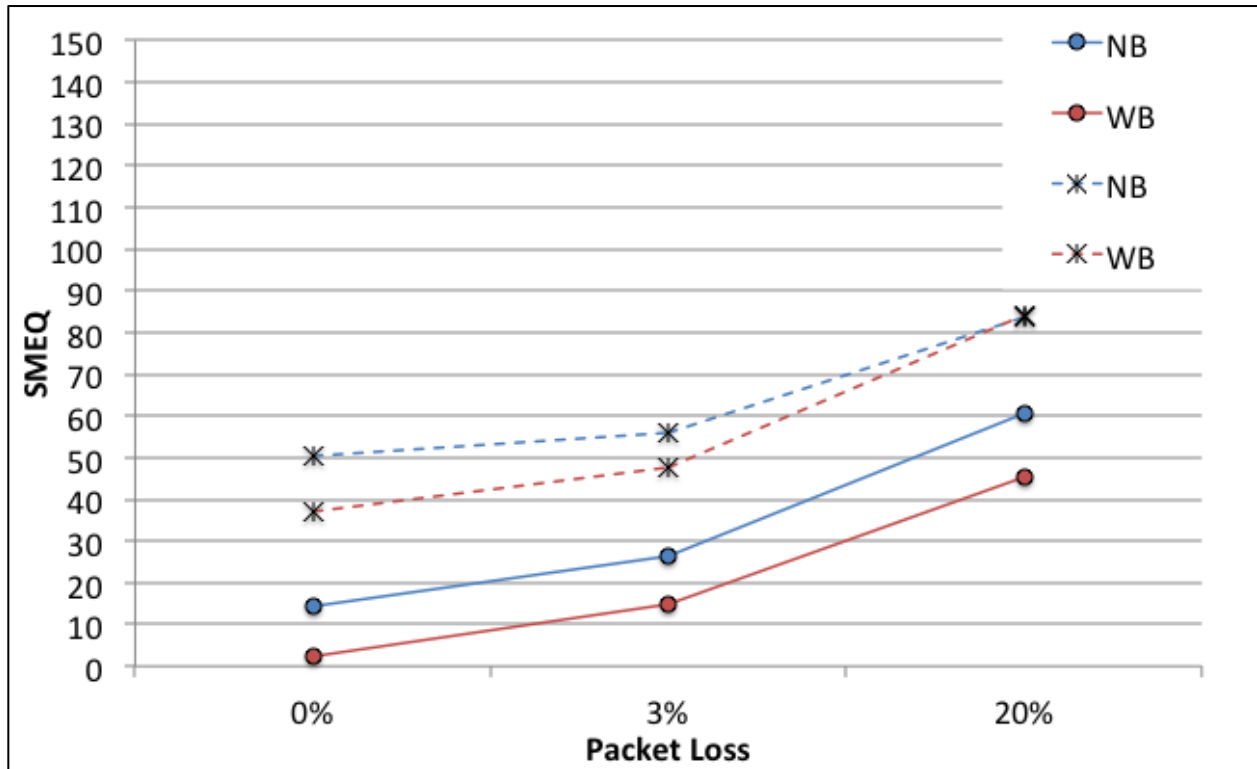


Figure 7: Subjective Mental Effort for Hearing Participants and Participants with Hearing Loss



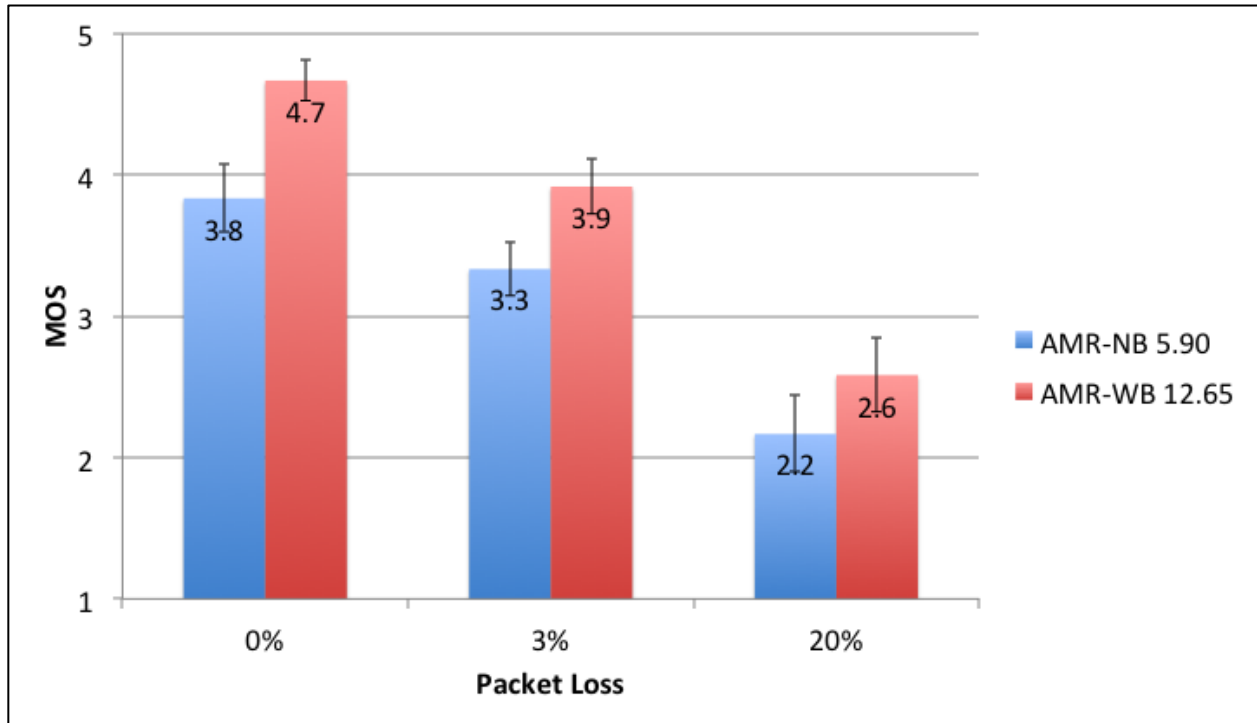
Mean Opinion Score (MOS)

Hearing Participants

Hearing participants' ratings of speech quality were reduced for NB speech as compared to WB speech and as packet loss increased. WB speech with no packet loss was judged to have near excellent speech quality, while NB speech with 20% packet loss was judged to have poor speech quality (Figure 8).

For the hearing participants, a repeated measures, two-way ANOVA for MOS ratings showed significant ($\alpha=0.05$) main effects of the factors audio bandwidth ($F_{(1,11)} = 22.6$, $p<0.001$) and packet loss ($F_{(2,22)} = 77.0$, $p<0.000$), but no significant interaction between the two factors ($F_{(2,22)} = 1.06$, $p<0.363$).

Figure 8: Mean Opinion Score for Hearing Participants



Participants with Hearing Loss

Ratings of speech quality by participants with hearing loss reduced as packet loss increased from 0% to 3% for both NB and WB speech. At the highest level of packet loss, participants with hearing loss judged speech quality to be similarly poor for both NB and WB speech. Without any packet loss, these same participants reported that speech quality was good for WB speech and fair for NB speech (Figure 9). Overall, participants with hearing loss judged speech quality lower than the hearing participants, regardless of audio codec bandwidth and degree of packet loss (Figure 10).

For participants with hearing loss, a repeated measures, two-way ANOVA for MOS ratings showed significant ($\alpha=0.05$) main effects of the factors audio bandwidth ($F_{(1,35)} = 7.75, p<0.009$) and packet loss ($F_{(2,70)} = 58.2, p<0.000$) and a significant interaction between the two factors ($F_{(2,70)} = 3.22, p<0.046$).

Figure 9: Mean Opinion Score for Participants with Hearing Loss

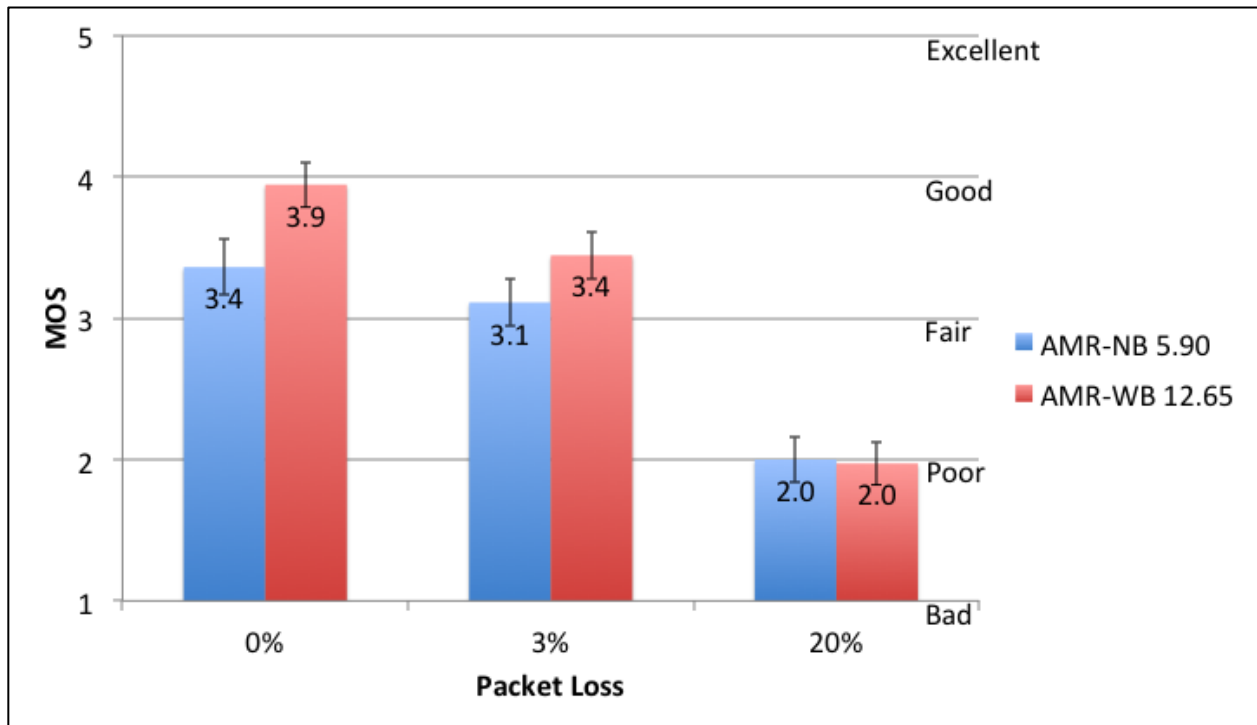
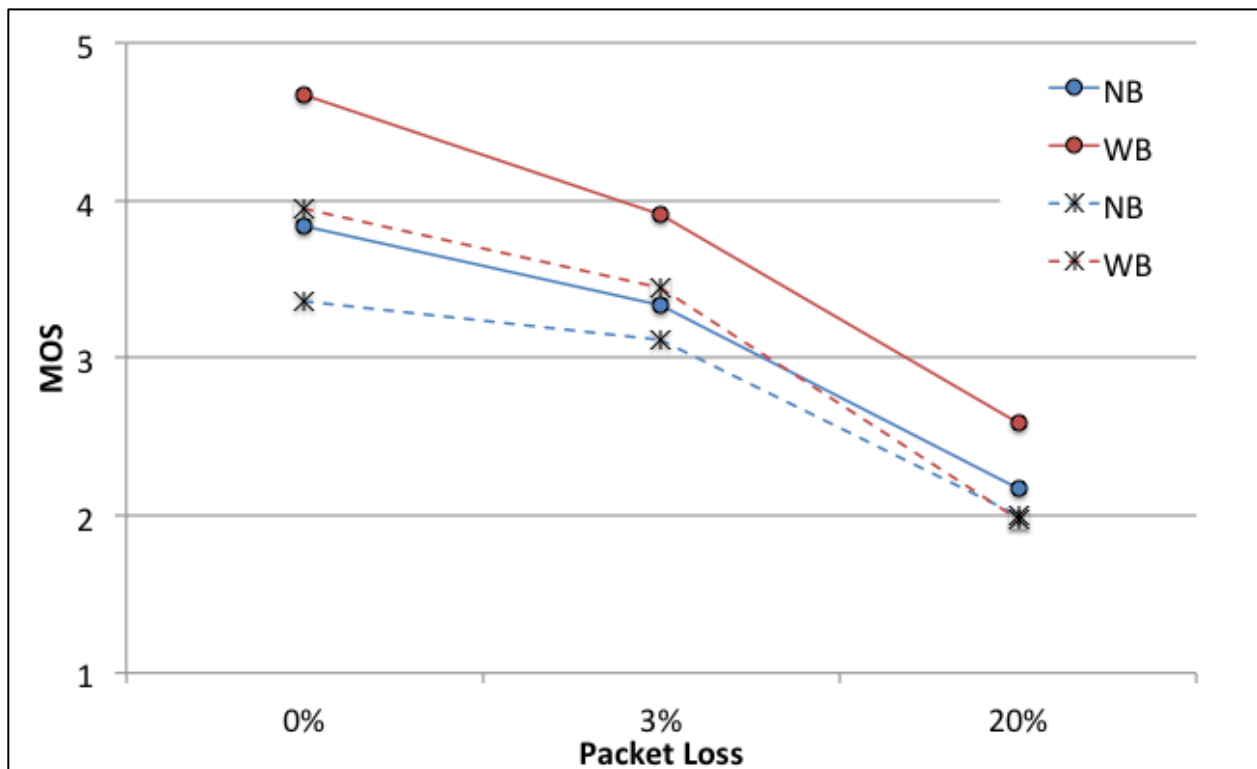


Figure 10: Mean Opinion Score for Hearing Participants and Participants with Hearing Loss



Purchase Intent

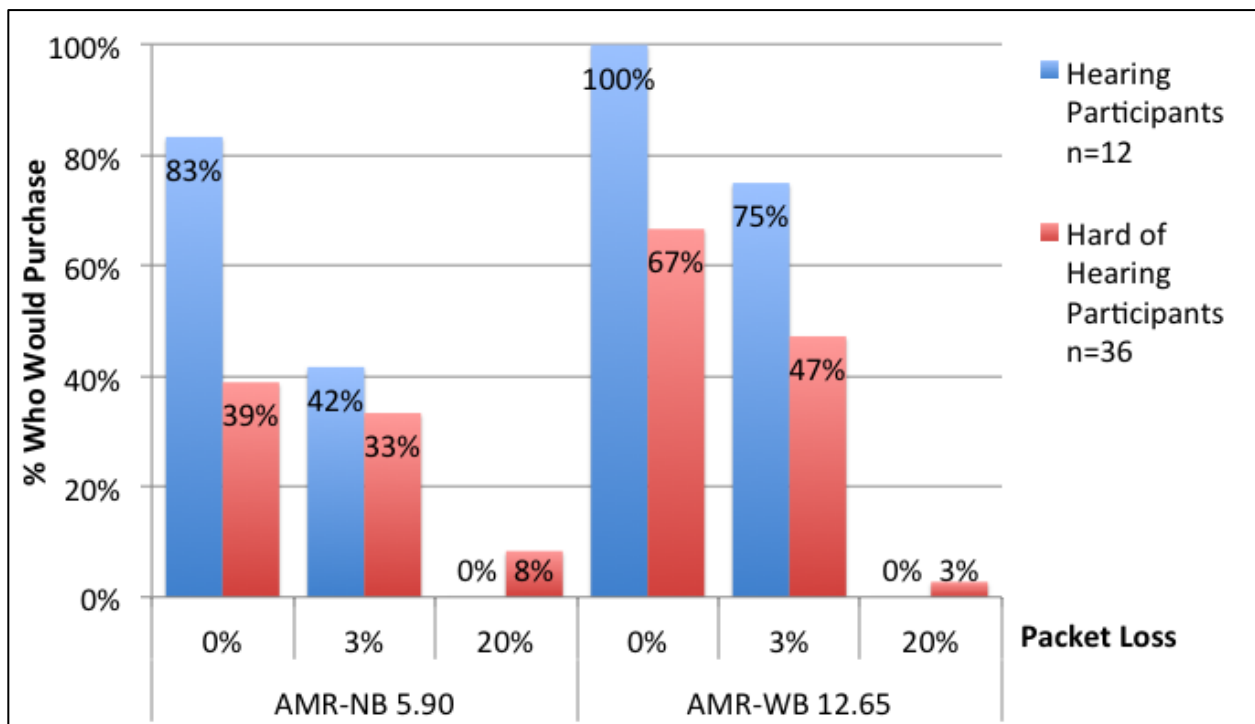
Hearing Participants

Purchase intent, based on listening experience alone, was higher for the hearing participants for WB than NB speech at both 0% and 3% packet loss. Intent to purchase reduced with increasing levels of packet loss until no hearing participants reported the intent to purchase at the highest level of packet loss regardless of audio bandwidth (Figure 11).

Participants with Hearing Loss

Participants with hearing loss had overall lower levels of purchase intent than the hearing participants, regardless of audio codec bandwidth and degree of packet loss. Intent to purchase was lower for NB speech than WB speech and reduced as packet loss increased (Figure 11).

Figure 11: Purchase Intent for Hearing Participants and Participants with Hearing Loss



For the paired nominal data from the purchase intent question, McNemar's Test was used to determine which pair-wise comparisons among all six conditions were significantly different from each other, for each participant group separately. A table summarizing the results can be found in Appendix D. Appendix D also contains a complete table of the inferential statistics for the other three dependent measures and all descriptive statistics. In all cases, data analyses were carried out separately for the hearing participants and those with hearing loss.

Conclusions

Overall, participants with hearing loss had poorer speech understanding, higher expenditures of mental effort, lower perceived speech quality and lower rates of purchase intent than hearing participants; regardless of audio bandwidth or degree of packet loss. Audio bandwidth and packet loss affected participants with hearing loss and hearing participants in similar ways.

Wideband audio provided an advantage in all dependent measures for both hearing participants and participants with hearing loss. However, among hearing participants, a wideband advantage for speech understanding only occurred at the highest level of packet loss. Otherwise, speech understanding for this group was the similarly high for narrowband and wideband audio. The participants with hearing loss showed a wideband advantage for speech understanding, albeit small, regardless of degree of packet loss. In previous experiments a larger advantage for wideband audio was found than here. However, in this study the sentence material was much more challenging than the materials employed in previous experiments. Additionally, speakers were changed within

single sentence sets – from the participants’ perspective seemingly at random, which may have made the task particularly difficult for individuals with hearing loss, and may have represented worst-case scenarios that rarely occur in practice.

For participants with hearing loss, the wideband audio advantage for perceived speech quality reduced as packet loss increased, which did not occur for the hearing participants. Both hearing participants and participants with hearing loss were least likely to purchase at the highest level of packet loss, regardless of audio bandwidth. At 0% packet loss, audio bandwidth did affect purchase intent for participants with hearing loss, but not for hearing participants.

As packet loss increased, performance decreased on all dependent measures for both groups. However, performance degraded more for participants with hearing loss than for hearing participants. Twenty percent bursty packet loss reduced performance to such a low level for individuals with hearing loss that they would likely not be able to use the voice telephone under these conditions regardless of the audio bandwidth available. This was not true for hearing participants; while performance would be degraded with 20% bursty packet loss, particularly under narrowband audio conditions, they would likely still be able to use the voice telephone. These results are consistent with well-documented findings of the greater susceptibility of individuals with hearing loss to considerable reductions in speech communication ability in other adverse listening situations, such as environments with competing noise and reverberation. Even so, the benefits of wideband audio for people with hearing loss observed in previous studies have been replicated in the present study, despite the employment of much more challenging stimuli. Using wideband audio may lead to more accessible voice communications in mobile telephony

environments for individuals with hearing loss whose peripheral auditory systems and hearing devices provide access to the increased frequency range afforded by WB audio.

Refining our understanding of the upper limits of a WB audio advantage under packet loss conditions and exploring other types and levels of network impairments are possible future research directions. Furthermore, extending this receive-only testing to conversational evaluations could lead to a better understanding of how these findings translate into real world improvements in voice telecommunications accessibility for individuals with hearing loss.

Acknowledgments

Laurie Garrison of AT&T Labs carried out all signal processing for speech coding and injection of packet loss.

Norman Williams developed the software for presenting the stimuli and recording participant responses.

Rosemary Johnson assisted in preparing the stimuli for encoding.

Appendix A

Sentence Set 1

Corresponds to Harvard Lists 11 and 14.

1. Oak is strong and also gives shade.
2. Cats and dogs each hate the other.
3. The pipe began to rust while new.
4. Open the crate but don't break the glass.
5. Add the sum to the product of these three.
6. Thieves who rob friends deserve jail.
7. The ripe taste of cheese improves with age.
8. Act on these orders with great speed.
9. The hog crawled under the high fence.
10. Move the vat over the hot fire.
11. A cramp is no small danger on a swim.
12. He said the same phrase thirty times.
13. Pluck the bright rose without leaves.
14. Two plus seven is less than ten.
15. The glow deepened in the eyes of the sweet girl.
16. Bring your problems to the wise chief.
17. Write a fond note to the friend you cherish.
18. Clothes and lodging are free to new men.
19. We frown when events take a bad turn.
20. Port is a strong wine with a smoky taste.

Sentence Set 2

Corresponds to Harvard Lists 15 and 16.

1. The young kid jumped the rusty gate.
2. Guess the results from the first scores.
3. A salt pickle tastes fine with ham.
4. The just claim got the right verdict.
5. These thistles bend in a high wind.
6. Pure bred poodles have curls.
7. The tree top waved in a graceful way.
8. The spot on the blotter was made by green ink.
9. Mud was spattered on the front of his white shirt.
10. The cigar burned a hole in the desk top.
11. The empty flask stood on the tin tray.

12. A speedy man can beat this track mark.
13. He broke a new shoelace that day.
14. The coffee stand is too high for the couch.
15. The urge to write short stories is rare.
16. The pencils have all been used.
17. The pirates seized the crew of the lost ship.
18. We tried to replace the coin but failed.
19. She sewed the torn coat quite neatly.
20. The sofa cushion is red and of light weight.

Sentence Set 3

Corresponds to Harvard Lists 17 and 19.

1. The jacket hung on the back of the wide chair.
2. At that high level the air is pure.
3. Drop the two when you add the figures.
4. A filing case is now hard to buy.
5. An abrupt start does not win the prize.
6. Wood is best for making toys and blocks.
7. The office paint was a dull sad tan.
8. He knew the skill of the great young actress.
9. A rag will soak up spilled water.
10. A shower of dirt fell from the hot pipes.
11. Acid burns holes in wool cloth.
12. Fairy tales should be fun to write.
13. Eight miles of woodland burned to waste.
14. The third act was dull and tired the players.
15. A young child should not suffer fright.
16. Add the column and put the sum here.
17. We admire and love a good cook.
18. There the flood mark is ten inches.
19. He carved a head from the round block of marble.
20. She has a smart way of wearing clothes.

Sentence Set 4

Corresponds to Harvard Lists 21 and 22.

1. The brown house was on fire to the attic.
2. The lure is used to catch trout and flounder.
3. Float the soap on top of the bath water.
4. A blue crane is a tall wading bird.

5. A fresh start will work such wonders.
6. The club rented the rink for the fifth night.
7. After the dance they went straight home.
8. The hostess taught the new maid to serve.
9. He wrote his last novel there at the inn.
10. Even the worst will beat his low score.
11. The cement had dried when he moved it.
12. The loss of the second ship was hard to take.
13. The fly made its way along the wall.
14. Do that with a wooden stick.
15. Live wires should be kept covered.
16. The large house had hot water taps.
17. It is hard to erase blue or red ink.
18. Write at once or you may forget it.
19. The doorknob was made of bright clean brass.
20. The wreck occurred by the bank on Main Street.

Sentence Set 5

Corresponds to Harvard Lists 23 and 25.

1. A pencil with black lead writes best.
2. Coax a young calf to drink from a bucket.
3. Schools for ladies teach charm and grace.
4. The lamp shone with a steady green flame.
5. They took the axe and the saw to the forest.
6. The ancient coin was quite dull and worn.
7. The shaky barn fell with a loud crash.
8. Jazz and swing fans like fast music.
9. Rake the rubbish up and then burn it.
10. Slash the gold cloth into fine ribbons.
11. On the islands the sea breeze is soft and mild.
12. The play began as soon as we sat down.
13. This will lead the world to more sound and fury
14. Add salt before you fry the egg.
15. The rush for funds reached its peak Tuesday.
16. The birch looked stark white and lonesome.
17. The box is held by a bright red snapper.
18. To make pure ice, you freeze water.
19. The first worm gets snapped early.
20. Jump the fence and hurry up the bank.

Sentence Set 6

Corresponds to Harvard Lists 26 and 29.

1. Yell and clap as the curtain slides back.
2. They are men who walk the middle of the road.
3. Both brothers wear the same size.
4. In some form or other we need fun.
5. The prince ordered his head chopped off.
6. The houses are built of red clay bricks.
7. Ducks fly north but lack a compass.
8. Fruit flavors are used in fizz drinks.
9. These pills do less good than others.
10. Canned pears lack full flavor.
11. The shelves were bare of both jam or crackers.
12. A joy to every child is the swan boat.
13. All sat frozen and watched the screen.
14. A cloud of dust stung his tender eyes.
15. To reach the end he needs much courage.
16. Shape the clay gently into block form.
17. The ridge on a smooth surface is a bump or flaw.
18. Hedge apples may stain your hands green.
19. Quench your thirst, then eat the crackers.
20. Tight curls get limp on rainy days.

Appendix B
Performance of AMRWB and NB in Packet Loss Conditions
VQA Test Plan
Laurie Garrison
May 27, 2015

Source Stimuli

Delivered by Gallaudet: 16bit, 16KHZ intel byte ordered with a 44 byte wav header, 72 total files

Test Conditions

1. Codecs: AMRNB rate 5.90 and AMRWB rate 12.65, DTX disabled for both
2. Packet loss conditions: 3% and 20%; gamma set to 0.8 for both
3. Baseline condition: no packet loss (0%)

Preprocessing Steps (all)

1. Semi-automatically trim silence from the beginning and end of audio files, one per sentence and speaker, with no padding, verified by a human listener
2. Save each trimmed file as standard wav file, 16 kHz PCM

Preprocessing Step (for each condition/sentence set combination)

For each participant from 1-12 do:

For each condition from 1-6 do:

Concatenate 20 trimmed sentences chosen for participant and condition into a single wav file with 1s of silence separating them

Output: 72 sentence files encoded as 16 kHz PCM, one per subject and condition (12 x 6)

Processing Steps- AMR-WB

1. Strip 44 byte wav header
2. Prefilter with ITU P.341 filter
3. Set level to -20dBm using ITU set_splevel
4. Process 16KHZ files through AMRWB coder, error-insertion-device as required using Gilbert-Elliot model
 21. Filename convention: talker-codec-all-processed

Processing Steps- AMR-NB

1. Strip 44 byte wav header
2. Prefilter with ITU P.341 filter
3. Set level to -20dBm using ITU set_splevel
4. Downsample to 8K using ITU filter HQ2 2:1
5. Swap bytes for processing on Sun
6. Process 8KHZ files through AMRNB coder, EID as required using Gilbert-Elliot model
 - a. Same filename convention as WB
7. Swap bytes back to little endian (Intel)

Total number of processed files to be delivered by ATT

72 files. Each file categorized by condition

AMRWB files will be raw, headerless, 16KHZ, 16bit Intel byte-ordered (little endian)

AMRNB files will be raw, headerless, 8KHZ, 16bit Intel byte-ordered (little endian)

Postprocessing steps (all)

1. Add WAV header back to raw files
2. Separate concatenated files back into constituent sentences and pad beginning and end of sentences with 100 ms of silence
3. Generate scripts to run through proper ordering of sentences and conditions for each subject

Appendix C

The dependent measures were speech understanding (% words correct), speech quality (mean opinion score – MOS), subjective mental effort (subjective mental effort questionnaire – SMEQ), and purchase intent. This appendix lists the materials that were used to elicit each measure.

Speech Understanding

See Appendix A for the sentence list.

Speech Quality

Participants were shown a screen with the following text and asked to rate the speech quality by selecting the appropriate option:

In this experiment, we are evaluating systems that might be used for voice telecommunications services. You are going to hear a number of recorded sentences. We would like you to rate how good they sound. You will use the following scale to provide your opinion of their overall quality.

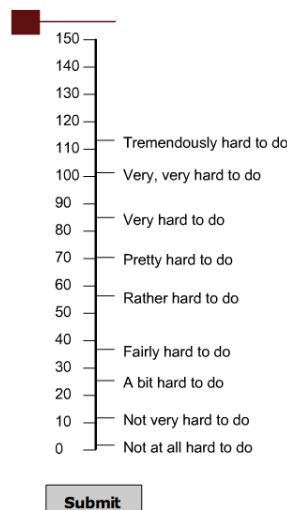
The overall quality of the speech was:

- Excellent
- Good
- Fair
- Poor
- Bad

Subjective Mental Effort

Participants were shown a screen with the following text and asked to rate the difficulty of the task on a slider:

How much effort did it take to understand what the person on the cell phone was saying?



Purchase Intent

Participants were shown a screen with the following text and asked to pick the appropriate option:

Would you purchase (and use) a cell phone with this level of speech quality?

Yes

No

Appendix D

Data Table for Hearing Participants

Condition	Hearing n=12	% Words Understood	SMEQ	MOS	Purchase (Yes/No)
NB 5.90 0%	Mean	98%	14.3	3.8	10/2
	SE	0.8%	3.1	0.2	
NB 5.90 3%	Mean	95%	26.2	3.3	5/7
	SE	1.1%	3.9	0.2	
NB 5.90 20%	Mean	83%	60.6	2.2	0/12
	SE	2.6%	7.1	0.3	
WB 12.65 0%	Mean	99%	2.3	4.7	12/0
	SE	0.9%	0.9	0.1	
WB 12.65 3%	Mean	99%	14.8	3.9	9/3
	SE	0.9%	3.0	0.2	
WB 12.65 20%	Mean	91%	45.2	2.6	0/12
	SE	1.7%	5.3	0.3	

Data Table for Participants with Hearing Loss

Condition	Hearing Loss N=36	% Words Understood	SMEQ	MOS	Purchase (Yes/No)
NB 5.90 0%	Mean	77.6%	50.3	3.4	14/22
	SE	3.5%	6.4	0.2	
NB 5.90 3%	Mean	73.6%	55.8	3.1	12/24
	SE	3.3%	5.4	0.2	
NB 5.90 20%	Mean	42.9%	83.8	2.0	3/33
	SE	3.1%	5.8	0.2	
WB 12.65 0%	Mean	81.2%	37.1	3.9	24/12
	SE	3.4%	5.2	0.2	
WB 12.65 3%	Mean	77.3%	47.6	3.4	17/19
	SE	3.5%	5.7	0.2	
WB 12.65 20%	Mean	51.2%	84.1	2.0	1/35
	SE	3.8%	5.3	0.2	

ANOVA: Two Within-Subject Factors

	Hearing Participants n=12	Participants with Hearing Loss n=36
Words Correct	<p>audio bandwidth $F(1,11) = 16.0$ $p < 0.002$</p> <p>packet loss $F(2,22) = 31.9$ $p < 0.000$</p> <p>audio bandwidth x packet loss $F(2,22) = 5.82$ $p < 0.009$</p>	<p>audio bandwidth $F(1,35) = 16.8$ $p < 0.000$</p> <p>packet loss $F(2,70) = 278$ $p < 0.000$</p> <p>audio bandwidth x packet loss $F(2,70) = 1.44$ $p < 0.243$</p>
SMEQ	<p>audio bandwidth $F(1,11) = 14.0$ $p < 0.003$</p> <p>packet loss $F(2,22) = 47.3$ $p < 0.000$</p> <p>audio bandwidth x packet loss $F(2,22) = 0.329$ $p < 0.723$</p>	<p>audio bandwidth $F(1,35) = 4.44$ $p < 0.042$</p> <p>packet loss $F(2,70) = 54.1$ $p < 0.000$</p> <p>audio bandwidth x packet loss $F(2,70) = 1.75$ $p < 0.181$</p>
MOS	<p>audio bandwidth $F(1,11) = 22.6$ $p < 0.001$</p> <p>packet loss $F(2,22) = 77.0$ $p < 0.000$</p> <p>audio bandwidth x packet loss $F(2,22) = 1.06$ $p < 0.363$</p>	<p>audio bandwidth $F(1,35) = 7.57$ $p < 0.009$</p> <p>packet loss $F(2,70) = 58.2$ $p < 0.000$</p> <p>audio bandwidth x packet loss $F(2,70) = 3.22$ $p < 0.046$</p>

McNemar's Test: Purchase Intent

Hearing Loss n=36	NB-0%	NB-3%	NB-20%	WB-0%	WB-3%	WB-20%
NB-0%		0.08, p=0.77	6.67, p=0.01	5.79, p=0.02	0.36, p=0.55	11.08, p=0.00
NB-3%	0.08, p=0.77		5.82, p=0.02	6.72, p=0.01	1.23, p=0.27	9.09, p=0.00
NB-20%	6.67, p=0.01	5.82, p=0.02		19.05, p=0.00	10.56, p=0.00	0.50, p=0.48
WB-0%	5.79, p=0.02	6.72, p=0.01	19.05, p=0.00		2.77, p=0.10	21.04, p=0.00
WB-3%	0.36, p=0.55	1.23, p=0.27	10.56, p=0.00	2.77, p=0.10		14.06, p=0.00
WB-20%	11.08, p=0.00	9.09, p=0.00	0.50, p=0.48	21.04, p=0.00	14.06, p=0.00	

Hearing n=12	NB-0%	NB-3%	NB-20%	WB-0%	WB-3%	WB-20%
NB-0%		3.20, p=0.07	8.10, p=0.00	0.50, p=0.48	0.00, p=1.00	8.10, p=0.00
NB-3%	3.20, p=0.07		3.20, p=0.07	5.14, p=0.02	2.25, p=0.13	3.20, p=0.07
NB-20%	8.10, p=0.00	3.20, p=0.07		10.08, p=0.00	7.11, p=0.01	undefined
WB-0%	0.50, p=0.48	5.14, p=0.02	10.08, p=0.00		1.33, p=0.25	10.08, p=0.00
WB-3%	0.00, p=1.00	2.25, p=0.13	7.11, p=0.01	1.33, p=0.25		7.11, p=0.01
WB-20%	8.10, p=0.00	3.20, p=0.07	undefined	10.08, p=0.00	7.11, p=0.01	