Factors Affecting the Accessibility of Voice Telephony for People with Hearing Loss: Audio Encoding, Network Impairments, Video and Environmental Noise

LINDA KOZMA-SPYTEK and CHRISTIAN VOGLER, Gallaudet University, Washington DC, USA

This paper describes four studies with a total of 114 individuals with hearing loss and 12 hearing controls that investigate the impact of audio quality parameters on voice telecommunications. These studies were first informed by a survey of 439 individuals with hearing loss on their voice telecommunications experiences. While voice telephony was very important, with high usage of wireless mobile phones, respondents reported relatively low satisfaction with their hearing devices' performance for telephone listening, noting that improved telephone audio quality was a significant need. The studies cover three categories of audio quality parameters: (1) **narrowband (NB)** versus **wideband (WB)** audio; (2) encoding audio at varying bit rates, from typical rates used in today's mobile networks to the highest quality supported by these audio codecs; and (3) absence of packet loss to worst-case packet loss in both mobile and VoIP networks. Additionally, NB versus WB audio was tested in auditory-only and audiovisual presentation modes and in quiet and noisy environments. With WB audio in a quiet environment, individuals with hearing loss exhibited better speech recognition, expended less perceived mental effort, and rated speech quality higher than with NB audio. WB audio provided a greater benefit when listening alone than when the visual channel also was available. The noisy environment significantly degraded performance for both presentation modes, but particularly for listening alone. Bit rate affected speech recognition for NB audio, and speech quality ratings for both NB and WB audio. Packet loss affected all of speech recognition, mental effort, and speech quality ratings. WB versus NB audio also affected hearing individuals, especially under packet loss. These results are discussed in terms of the practical steps they suggest for the implementation of telecommunications systems and related technical standards and policy considerations to improve the accessibility of voice telephony for people with hearing loss.

CCS Concepts: • **Human-centered computing** → **Empirical studies in accessibility**; *Accessibility design and evaluation methods*;

Additional Key Words and Phrases: Voice telephony, telecommunications, hearing loss, hard of hearing, cochlear implant, hearing aid, wideband audio, narrowband audio, audio codec, bit rate, packet loss

This paper was developed under a grant from the AT&T Foundation, and under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90RE5003). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this paper do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government. Paula Tucker carried out many of the experimental procedures and data management. 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, data collection and participant interaction. Mary Garvert assisted in the data collection and participant interaction.

Authors' address: L. Kozma-Spytek and C. Vogler, Gallaudet University, 800 Florida Ave NE, TAP - SLCC 1116, Washington DC 20002, USA; emails: {linda.kozma-spytek, christian.vogler}@gallaudet.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1936-7228/2021/10-ART21 \$15.00

<https://doi.org/10.1145/3479160>

ACM Reference format:

Linda Kozma-Spytek and Christian Vogler. 2021. Factors Affecting the Accessibility of Voice Telephony for People with Hearing Loss: Audio Encoding, Network Impairments, Video and Environmental Noise. *ACM Trans. Access. Comput.* 14, 4, Article 21 (October 2021), 35 pages.
<https://doi.org/10.1145/3479160>

1 INTRODUCTION

Telephone communication for individuals with hearing loss has long been problematic for a variety of reasons. Hearing loss can moderately-to-greatly affect an individual's use of the telephone causing them to either be discouraged from or avoid use of the telephone. Problems include trouble hearing the telephone ring, an unwillingness to answer the telephone, and difficulty coupling a hearing device to the telephone and then listening to the conversation [61, 33]. These difficulties with telephone communication can affect both an individual's social life and job performance. One study [60] identified telephone communication as the most problematic type of interaction in the workplace for individuals with hearing loss, with some considering retirement because of the difficulties they experience using the telephone.

Fundamentally, speech understanding by people with hearing loss is significantly less robust [11] compared to that of hearing people, and telecommunications has been for the most part voice-only, without the benefit of the lipreading enhancement afforded by face-to-face communication [36]. More recently, cellphone use has moved telephone conversations from mostly private locations to also include public venues, such as on public transportation, in airports and at restaurants. The use of cellphones in public spaces means calls are taking place in more complex acoustic environments that may include noise on one or both ends of the call.

1.1 Accessibility for Voice Telecommunications

To improve phone call access for people with hearing loss, the United States Federal Communications Commission has established (1) a captioned telephone relay service and (2) **hearing aid compatibility (HAC)** rules for landline and wireless telephones. The captioned telephone relay service provides audio accompanied by captions generated by human operators. However, this popular service has been beset by increasing cost [15] due to rising growth in usage and does not offer the same degree of privacy and effective communication as direct voice calls that do not involve a third-party operator. It also is available in only a few countries worldwide (as of 2021, this includes the US, Australia, and New Zealand).

The HAC requirements have focused on optimizing delivery of the phone audio to a person's hearing device via the built-in telecoil coupling capability of handsets [35], the reduction of audible **radio frequency (RF)** interference noise in hearing devices from wireless phones, and the delivery of adequate audio volume levels from a telephone's receiver.

While these efforts have been important in making the telephone more accessible to people with hearing loss who use hearing devices and voice communication, these improvements have been driven by research findings and technology that date back more than a decade. Since then, the behavior of people with hearing loss has evolved, as has telecommunications technology and its impact on call quality.

1.2 Audio Quality on Phone Calls

In voice phone calls the audio quality of the transmitted speech is typically reduced compared to face-to-face communication. Moreover, with the significant move away from copper-wire landline telephones in favor of VoIP and wireless cellular telecommunications, speech quality over

Fig. 1. Narrowband versus wideband full-scale dB audio power spectrum from a sentence drawn from the IEEE Harvard sentence set: “Cats and dogs each hate the other.” There is approximately 20% more speech information available in the frequencies > 3400 Hz.

these newer networks has reduced. In mainstream mobile and VoIP telephony, ratings of network audio quality have consistently demonstrated poorer quality for speech than for that experienced on the analog **public switched telephone network (PSTN)**. These reductions in quality are due in part to technical parameters related to the audio encoding on telecommunications networks, including the audio bandwidth and data rates, as well as reductions in mobile and VoIP network performance. Audio encoding parameters and degraded network performance also may disproportionately impact the accessibility of telephone networks. For people with hearing loss, intelligibility, in addition to speech quality, may be negatively affected, and consequently, access to the telephone also may be impacted.

In terms of audio bandwidth, many phone networks do not transmit sounds outside approximately 300–3,400 Hz [28]. Some acoustic cues important for speech intelligibility are at these higher frequencies; access to them increases the amount of speech that may be available by approximately 20% [34] (Figure 1). The limitations of narrowband audio (also called standard definition voice) originally stem from the PSTN, which has long formed the backbone of telecommunications. Narrowband communication channels also offer significant advantages with respect to the amount of spectrum and transmit power needed in radio broadcasting applications [8].

Significant work has been done in the area of coding to optimize narrowband audio in digital transmission under constrained bandwidth. One area of improvement consists of allocating bits of information dynamically as a refinement to **linear predictive coding (LPC)** [57, 58]. Another improvement was to reduce bit rates via using a codebook in place of LPC [9]. Variable frame transmission also has been considered for reducing the bit rate [62].

If data bandwidth constraints can be relaxed, a simple way to improve sound quality, and also potentially improve access for people with hearing loss, is to use wideband encoding (also called **high-definition** or **HD voice**). Wideband audio codecs typically extend the audio bandwidth from the narrowband 300 Hz–3.4 kHz to 50 Hz–7 kHz, although the effective bandwidth will be circumscribed to some degree by other factors, such as the handset characteristics on both the send and receive ends. The **3rd Generation Partnership Project (3GPP)** standards group adopted the AMR-WB codec for implementing wireless cellular wideband audio. Additionally, wideband encoding techniques have been employed in point-to-point VoIP calls (e.g., Skype), business telephone systems, and videoconferencing systems.

Many people with hearing loss have access to the wideband audio frequencies above the narrowband cutoff of 3400 Hz. Among hearing devices used by individuals with hearing loss, all three

cochlear implant (CI) systems available in the US can process audio and deliver information from approximately 200 Hz up to 8.5 kHz [38] depending on the system. **Hearing aids (HA)** have, likewise, extended their effective audio bandwidth and are being fit for individuals who demonstrate sound audibility and benefit from access to this extended frequency region [3, 6, 55].

For hearing individuals, increasing the telephone audio bandwidth has been shown to increase speech quality and, in some cases, intelligibility in noise [17]. However, no such effect was found for identifying voices already known to hearing individuals [55]. For individuals with hearing loss, a few studies have explored artificial telephone bandwidth extension, showing some promise for improved speech recognition and quality, primarily for cochlear implant users. [25, 41, 49]. There also has been some limited work to understand the possible benefits of wideband encoding for individuals with hearing loss, but the results of the studies differed. In one study, the experimenters applied an HA fitting strategy to the NB and WB cellular phone encoded audio and delivered it via insert earphones rather than using the participants own HAs. No significant WB audio advantage was found for speech recognition or for subjective ratings of quality and listening effort in quiet or in noise [42]. This result differed from those of two studies carried out by a research group from the Inselspital in Bern, Switzerland, in which VoIP WB telephony was found to produce better speech recognition for both HA and CI users than either NB audio over the PSTN or NB audio over the GSM mobile network [18, 45]. These latter studies either did not address mobile wireless audio encoding directly or did not control for the codecs used. While the first study did both, the codecs used are not commonly employed in mobile wireless networks, and the coupling and hearing loss compensation methods used do not reflect real-world scenarios.

Like audio bandwidth, the data bandwidth of the speech encoded bitstream can vary to improve the efficiency of the voice transmission and meet the requirements of the end device. The most common codec used for mobile networks (AMR-NB) has significantly reduced the bit rate from the 64 kbps bit rate on VoIP networks (with quality comparable to PSTN) to rates of 5.90 to 12.20 kbps via lossy compression techniques based on perceptual coding. AMR-WB also supports a range of bit rates from 12.65 to 23.85 kbps. Low bit rate connections have benefits in terms of decreasing telecommunications costs and increasing service availability, but there is a speech quality trade-off. Relatedly, varying bit rates and codecs have been studied in CI users as a means to reduce bandwidth and required power for signal processing [22, 23, 14]. While there is some potential overlap with telephony in the use of specific codecs, most of this work is CI-specific. Lower bit rates, as measured either directly or through a predictive audio quality algorithm such as PESQ [28], ViSQOL [20] or POLQA [21, 20], are associated with lower perceived audio quality for hearing individuals [1, 24]. Although predictive algorithms have been tested on people with hearing loss [15], the impact of bit rates on such individuals has not been directly studied.

Both VoIP and wireless mobile telephone environments are susceptible to impairments in network performance. One common network impairment is packet loss. Packet loss occurs to varying degrees in these telephony environments due to lost or delayed packets during network transmission. While packet loss is known to degrade perceived audio quality for hearing individuals [1, 24], the impact of this common impairment to network performance for individuals with hearing loss has been examined in only a single study [44]. High bit-rate WB VoIP audio with varying degrees of packet loss was compared to NB audio found over the PSTN in quiet and in the presence of varying signal to noise ratios. Participants included CI users, HA wearers and hearing individuals. Packet loss reduced speech recognition, with the most severe level of packet loss (20%) producing levels of performance for WB audio that were equal to those of NB audio without packet loss. Noise further degraded performance. Interpretation of results was compromised by the relatively high number of comparisons (60) that were made for a small number of participants (10 per group).

There is also some past work on other technical parameters related to accessible telephone systems for people with hearing loss. Such work includes the impact of audio-video sync for video telecommunications on listening and lip-reading [36, 43], sign language video communications [8, 12, 57], signal to noise ratios for coupling digital wireless telephones with hearing devices [30, 31], and more generally, connectivity options between phones and hearing devices [46, 51, 64]. Other research has focused on building audio enhancements directly into telecommunications systems [32, 53], in order to match the hearing loss profiles of individuals with their use of telephones. However, to our knowledge, very limited work has been done to date on audio bandwidth and network packet loss, no work has been done on bit rate for individuals with hearing loss, and only one study looked specifically at wireless mobile telephony. This is in spite of the prevalence of hearing loss. In the United States, the prevalence of hearing loss is about 12.7% or 30 million people 12 years or older [39], increasing with age to nearly 63.1% of adults 70 years or older [40]. The four experiments described in Section 4 explore the impact of technical parameters related to audio encoding and network performance on wireless mobile telecommunications for people with hearing loss.

2 PAPER OVERVIEW AND ORGANIZATION

To better understand the current experience of people with hearing loss, we first conducted a survey, described in detail in the next section (Section 3). The results of the survey were subsequently used to inform, in part, the factors of interest in a series of iterative experiments described later in this paper. These four experiments investigate the impact of audio quality factors (Section 4) on the accessibility of phone calls by people with hearing loss. In Experiment 1 we explore audio bandwidth in an audio-only setting in quiet environments under optimal network and listening conditions. While the effect of adding a video channel, video frame rate, and audio-video synchrony on the ability of people with hearing loss to understand spoken language during video telephone conversations had been studied previously [36], it did not consider audio bandwidth. Likewise, noisy environments are common in mobile telephony, but the effect of audio bandwidth in such environments had not been previously studied. These considerations, along with the results of Experiment 1, led to testing audio bandwidth in conjunction with video and noisy listening environments in Experiment 2. The results suggested that further investigation into audio-only environments were warranted. To this end, Experiment 3 follows up on audio bandwidth in audio-only settings with varying bit rates. Finally, based on the bit rate findings, in Experiment 4 we explore network packet loss typical of wireless mobile networks. We provide an overall discussion of the findings, along with recommendations, in Section 5. Note also that Appendix A contains a glossary of frequently used terms from the telecommunications field.

2.1 Expanded Conference Paper

This paper is an expanded version of a conference paper published in ASSETS 2019 [37]. The earlier paper reported on Experiments 1, 3 and 4 (in Sections 4.1, 4.3, 4.4), and a more condensed version of the main findings, applications and limitations (in Section 5). In this expanded version, we provide detailed results from a survey of people with hearing loss on their voice telecommunication behavior and experiences (in an all-new Section 3) that helped to motivate and frame the series of studies covered in this paper. We also extend the information on the factors of audio bandwidth, packet loss, and codecs from the conference paper by adding a new experiment to evaluate the effects of including a video signal to accompany the audio signal and the presence of environmental noise during telecommunications, for both narrowband and wideband audio (in a new Section 4.2). We further provide additional detail on the experimental procedure to make it easier to replicate, including more information on the technical setup (across all methods and materials subsections

of Section 4), as well as an appendix that describes the modeling used to implement packet loss (Appendix B).

3 TELECOM SURVEY OF PEOPLE WITH HEARING LOSS

We now describe selected results for a survey we conducted on voice telecommunications by adults who are hard of hearing, deaf or have hearing loss. We conducted the survey to learn about their behaviors and experiences in the use of voice telecommunications technology. Our goal was to better understand how such adults use current voice telecommunications technology, what barriers they face using it, and what needs they have for improved accessibility.

We present these results in aggregate to understand the overall experience of individuals with hearing loss. It is well known that the perceptual consequences of hearing loss and hearing device outcomes are highly varied. Additionally, while pure-tone audiometry is a dominant metric in characterizing hearing loss, it fails to fully account for the auditory experiences of people with hearing loss or explain their speech perception abilities, particularly in complex acoustic environments. Neither can it predict their hearing device outcomes, irrespective of what type of hearing device is used. Therefore, the information we provide on self-reported degree of hearing loss and hearing device use is meant to fully describe the range of individuals who responded to our survey but is not used to further differentiate the experiences of our respondents in their regular use of voice telephony, or to divide participants into different categories in the subsequent audio evaluation experiments.

3.1 Survey Methods

In the survey, individuals were asked questions about their hearing loss, hearing device use, the telephone technology they use, and the barriers they face and needs they have in the use of the telephone. To take this survey, respondents were required to be an adult (18 years or older) and have a hearing loss. Respondents were also screened for regular use of voice telecommunications (at least once a week). For the purposes of this survey, voice telecommunications was defined for respondents in the following way: “you *both listen and talk* for yourself during telephone calls, even if you also supplement your listening experience by using text (for example, relay or captioning) to read what the other person on the call is saying while you listen.” This anonymous, web-based, convenience survey was open for approximately one year, with active recruitment solicited twice during this time through national consumer organizations. A total of 439 respondents met all the survey inclusion criteria.

3.1.1 Survey Respondents. Approximately one-third of respondents (32%) were 60–69 years old; a quarter (25%) were 70+ years old and the remainder (42%) were between 18 and 59 years old (Figure 2). A little more than two-thirds (69%) identified as female, while the rest (31%) identified as male. Many (59%) lived in an urban locale, and the remainder lived in small cities or towns or in rural locales. A little more than two-thirds of respondents (68%) lived in households of two or more people; 32% lived alone. Most respondents (81%) had a college degree, with 44% working, 44% retired, and the rest students, homemakers or unemployed.

Self-reported degree of hearing loss ranged from mild to profound, with most respondents having moderately severe to profound hearing loss (Figure 3). Of the 439 respondents, 393 (90%) used a hearing device at the time of the survey while 46 did not.

Most hearing device users (80%) wore hearing aids, with some 20% wearing cochlear implants or in a few cases, another type of hearing device. These mostly experienced hearing device wearers had more than three years of hearing device use (91%) and wore their devices seven or more hours a day (88%). Most respondents (77%) reported having daily voice telephone calls, with a little more than half (51%) reporting three or more voice telephone calls a day (Figure 4).

Fig. 2. Percent of survey respondents (n = 439) by age given in decades.

Fig. 3. Number of survey respondents reporting the degree of hearing loss for each ear.

3.2 Survey Results

A little less than a quarter (23%) of survey respondents reported being either very satisfied or satisfied with the performance of their hearing device when using it for telephone listening. Most (61%) were somewhat satisfied, neutral or somewhat dissatisfied with their hearing device performance for telephone listening. The remaining 17% were dissatisfied or very dissatisfied (Figure 5).

A volume control (71%) and telecoil (64%) were the most common features hearing devices had that respondents reported were useful for telephone listening. One-third (33%) reported having a streamer with Bluetooth wireless technology and less than 20% reported having either feedback control for microphone listening or a program specifically for telephone listening. While these features were considered important for successful use of the telephone and, in the case of feedback control, very important, respondents were only somewhat satisfied with the performance of these features when they used the telephone.

Respondents reported using a variety of phone types in their daily lives, including land-line/Internet VoIP phones, cell phones and specialty phones. A specialty phone is a corded or cordless phone that provides additional amplification (i.e., higher gain than required by FCC hearing

Fig. 4. The percent of survey respondents and the frequency with which they reported voice telephone calls.

Fig. 5. Percent of survey respondents and their level of satisfaction with hearing device performance for telephone listening.

aid compatibility rules), tone control (i.e., a user adjustment to make particular frequency regions in the telephone audio signal softer or louder), and/or captions of a calling partner’s speech.

Most respondents (86%) owned a cell phone and for nearly half of them (48%), their cell phone was the phone they used most often for voice telecommunications. A majority of cell phone users (75%) owned a smartphone. In spite of the high rate of cell phone usage, only a little more than a third (37%) were satisfied with their calling experience on their personal cell phone. Cell phone users were most likely to hold the phone to their ear and use their hearing device’s microphone for listening to their calling partners.

In fact, many respondents (70%) reported holding the phone to their ear and using their hearing device as the most common method employed for telephone listening. Many fewer respondents (12%) reported using an accessory to connect the phone to their hearing device. At the time of the survey, directly connecting a phone and hearing device via Bluetooth was not possible. With this newer method, now available in a limited number of higher-end hearing devices, the phone does not need to be held to the ear and no additional hardware accessories are required for access to the audio signal of the phone.

Across the various telephones used most often, hearing device wearers (43%) reported using microphone coupling of some sort more than any other coupling method. In the case of microphone coupling, the acoustic signal produced by the telephone’s receiver is picked up by the microphone

Fig. 6. Percent of survey respondents and the level of speech understanding they reported during face-to-face communication in quiet and noisy environments.

of the hearing device, and little if any set up is required by the hearing device wearer for placing or receiving calls.

Some 64% of hearing device wearers reported having a telecoil in their hearing device, with only one-third (34%) of those with telecoils commonly using them for telecommunications. For this inductive form of coupling, the telecoil in the hearing device picks up a magnetic signal produced from the electrical signal that drives the telephone's receiver. And while one-third (33%) of hearing device wearers had a streamer with Bluetooth wireless technology, where the phone couples with the streamer via Bluetooth, and the streamer in turn couples with the hearing device, only a quarter (23%) of those with streamers used them as their most common method of telephone coupling. All other respondents most often removed their hearing device, did not use a hearing device at all, or used some other form of coupling.

During face-to-face voice communication in a quiet environment, many of the respondents (72%) reported being able to understand all to most of what their conversational partner was saying. Speech understanding reduced during face-to-face communication in the presence of noise and during telephone communication (Figure 6). During face-to-face communication in noise, only 14% of respondents reported being able to understand all to most of what their partner was saying. On the telephone, the percentage of respondents who reported being able to understand all to most of what their calling partner was saying dropped to 50% (Figure 7).

Other than the speech characteristics of their calling partner and the individual's own hearing loss, telephone sound quality and environmental noise during phone use were reported as two of the biggest barriers to satisfactory telephone communication. Correspondingly, the most important need these respondents identified in order to achieve successful telephone communication was better telephone sound quality [47].

Finally, specific comments we received from individual survey respondents indicate that telephone communication is challenging at best, particularly for those who want to both listen and speak for themselves. They report that when they cannot make out individual words and hence fail to piece together conversations through listening alone, that it is frustrating, tiring, stressful

Fig. 7. Percent of survey respondents and the level of speech understanding they reported during face-to-face communication and telephone communication.

and embarrassing. In spite of this, voice telephone use remains very important. Many individuals report the need for better sound quality and better clarity particularly on cell phones, emphasizing that simply making speech louder is not enough.

3.3 Survey Discussion

These survey results suggest that improving the accessibility of voice calls for people with hearing loss would be high-impact and could be cost-effective if it reduced the use of captioned telephone relay service. It is clear from the survey results that phone calls are important to people with hearing loss, with many of them placing multiple calls per day. Despite the frequency at which people make and receive calls, less than a quarter of all respondents report being satisfied with their hearing device for phone calls, leaving room for improvement. There was a high rate of cell phone ownership consistent with the rate of cell phone ownership found more generally in the United States [51]. Cell phones were also most often the phone of choice for making personal phone calls, even though more than 60% of those were less than satisfied with the listening experience. The survey results underscore the importance of speech quality in achieving satisfactory telephone communications. Hearing device features useful for telephone communication, while important, cannot overcome problems related to the inherent quality of the audio signal received by the hearing device. Given the high rate of ownership and usage of cell phones, investigating audio quality parameters that affect mobile wireless calling would be of particular importance. Additionally, respondents identified environmental noise, common in mobile telephony, as a major barrier to telecommunications, thereby making the study of both audio quality and its interaction with noise two promising avenues for improving satisfaction levels and potentially accessibility.

4 EVALUATION OF AUDIO QUALITY PARAMETERS

We conducted a series of four related experiments on voice telecommunications accessibility for individuals with hearing loss over the course of three years. The four experiments followed an iterative approach in which each successive experiment addressed a separate research question

while also building on questions raised by the findings of the previous experiments, replicating certain key findings with a new group of participants and strengthening our experimental protocol. The audio quality parameters investigated included codec audio bandwidth (i.e., narrowband vs wideband audio), codec bit rate, and the impact of packet loss. The core factor common across all experiments was the codec audio bandwidth – narrowband vs wideband audio. As such, we do not differentiate our results on the basis of either hearing device use or self-reported degree of hearing loss. Rather, we aggregate the results of all participants with hearing loss together for each experiment because the more important attribute was their access to WB audio through their hearing devices, which they all have in common.

Overall, the experiments address the following research questions for people with hearing loss in each respective experiment:

- Experiment 1: Does wideband (WB) audio increase speech recognition and decrease mental effort compared to narrowband (NB) audio?
- Experiment 2: Does any advantage of WB audio over NB audio hold when lipreading information is available through videotelephony and when the acoustic environment is noisy?
- Experiment 3: Do the data rates (i.e., bit rate) of the NB and WB audio codecs impact speech recognition and quality, and how do they compare to the quality of NB audio on the PSTN?
- Experiment 4: What is the impact of packet loss, characteristic of mobile and IP networks, on speech recognition, subjective mental effort and speech quality ratings; and how does any impact experienced by people with hearing loss compare to that experienced by hearing people?

Each experiment used a within-subjects, repeated measures design. Because each participant receives all experimental conditions with this type of design, each individual acts as his/her own control, and the variability introduced by individual differences among participants can be effectively addressed. During the one-hour test sessions, the paid participants listened to stimuli (available at <https://bit.ly/2JLF9Pj>) and provided speech recognition and ratings data.

For each experiment, participant recruitment took place through Hearing Loss Association of America and other institutions serving people with hearing loss. Hearing participants were recruited at Gallaudet University. All participants were required to be fluent English-speaking adults, 18 years of age or older. Participants with hearing loss were also required to be daily hearing device users, as well as regular users of the voice telephone. Additionally, they completed an intake survey at the beginning of their test session, which included questions about their degree of hearing loss, hearing device use, and communication during face-to-face and telephone conversations. This intake survey had some questions similar to the ones in the survey described in Section 3, but was otherwise distinct.

For all participants, we ensured that they were able to access the higher frequencies provided by wideband audio, as described in the following. In the first two experiments, only daily CI users were recruited, who are known to perceive sound up to at least 7 kHz [38], and thus are guaranteed to perceive the additional information included in wideband audio. In the last two experiments, both daily HA and CI users were recruited. All participants for these last two experiments, including the hearing participants, were required to pass a hearing screening for audibility of the higher frequencies (4 kHz and 5 kHz) available through wideband audio codecs. The screening verified each participant's ability to hear third-octave bands of noise centered at 4kHz and 5kHz played on the same signal presentation equipment and at the same levels as the mean level of these frequency regions in the speech stimuli used for the experiment (see Figure 1 for the approximate levels). For participants with hearing loss, the screening was completed while using their hearing devices.

In the subsequent subsections, details of the methods and results are described in turn, followed by a discussion of the findings for each experiment.

4.1 Experiment 1: Codec Audio Bandwidth

The research question addressed in Experiment 1 was: Does the use of wideband (WB) encoding compared to narrowband (NB) encoding increase speech recognition and decrease expenditures of mental effort for individuals with hearing loss who have access, through their hearing devices, to frequency information above 3,400 Hz?

4.1.1 Participants. Testing of NB versus WB telephone speech was completed with a group of 42 CI users, who as noted in the section overview, can access the frequencies above the NB cutoff of 3,400 Hz. Of these 42 individuals with hearing loss, 29 were women and 13 were men, with an average age of 57.5 years (ranging from 22–86 years). All participants had at least two to three years of self-reported hearing device use. All individuals used their cochlear implants during testing, with 23 bilateral CI users, 8 unilateral CI users and 11 individuals who used a CI in one ear and a HA in the other ear. Self-reported hearing loss ranged from severe to profound across both ears. Most participants (37) reported profound hearing loss in the ear(s) with a CI.

4.1.2 Materials. Stimuli for the experiment were drawn from the **Computer Assisted Speech Perception Evaluation and Training** tool, or **CASPER** [4]. CASPER is a system for evaluation of speech in audio, visual (lipreading) and combined modes that consist of 72 sentence sets with 102 words each. All CASPER sentences are spoken by one male and one female speaker, who are fluent, native English speakers with typical speech and no discernable accents. These sets consist of related topics and are designed to be representative of conversational speech. Eight sentence sets were used to prepare the stimuli for both test conditions.

Stimulus preparation: For each study participant, speech recognition was tested using CASPER sentence sets spoken by the female speaker in two audio-only conditions, narrowband telephone speech (AMR NB at a bit rate of 12.20 kbit/s) and wideband telephone speech (AMR WB at a bit rate of 23.85 kbit/s), which are the respective highest-quality rates for both codecs.

The original CASPER files are encoded as **Pulse Code Modulated (PCM)**, see also the glossary) audio sampled at 22,050 Hz. The audio stream was demultiplexed and downsampled to PCM audio at a sampling rate of 16 kHz, thereby limiting the audio bandwidth to 8 kHz. Then each stimulus was transcoded (see also the glossary) to AMR-NB and AMR-WB via *ffmpeg* with the OpenCore-AMR libraries [50] at the respective bit rates. Because the resulting samples are not widely supported by operating system playback and audio editing software, each AMR-encoded stimulus was losslessly converted back to PCM audio, again using *ffmpeg* with the OpenCore-AMR libraries. Finally, in order to ensure consistent playback levels across all stimuli, each file was level-equalized without affecting its dynamics or frequency content in a post-processing step in Adobe Audition.

All sentence sets were processed for each of the two test conditions. The sentence sets used for each condition were then counterbalanced across subjects. This was done to guard against the effects of possible differences in intelligibility, either inherent or as a result of the processing. Additionally, each participant received different sentence sets for the two test conditions to guard against learning effects.

4.1.3 Method. Participants repeated the sentences that they heard through the microphone of their CI using a test set up that simulated cell phone listening via speakerphone. Speakers were set up directly in front of the participant at 0° azimuth and close enough to be perceived as a single source, with a participant seated approximately 12 inches away from an imaginary base line connecting the two speakers. The volume was adjusted with an average presentation level set

Fig. 8. Experimental setup for narrowband vs wideband audio showing the speaker setup. The speakers were close enough to be perceived as a single source. While this was 16 inches of separation in our setup, different speakers may require a different amount of separation, and this perception needs to be verified by listening. Note that in this setup two speakers were used, because some unrelated conditions that were tested as part of the same experimental session required use of a screen in between. If no screen had been used, a single calibrated speaker placed directly in front of the participant would have been equivalent.

Table 1. Mean Percent Words Correct Statistics for Wideband WB x NB Audio (n = 42)

Codec	%Words Correct	Std Dev	Std Error
AMR-WB 23.85	75.7	23.4	3.6
AMR-NB 12.20	65.2	26.4	4.1

such that the frequent peaks of the speech stimuli measured between 64–70 dB SPL on a sound level meter at the location of the listener’s head (see Figure 8). This range was selected because it represents both the average level of conversational speech during typical face-to-face communication at one meter and the nominal levels for conversational gain used in telephony for one-ear and two-ear listening [7]. The participants’ responses were scored as the number of words correctly repeated per set (out of 102 total words) for each condition. Each administration of one sentence set per condition took approximately five minutes. Presentation of conditions was counterbalanced across subjects.

Following the completion of testing for one sentence set/one condition, the **Subjective Mental Effort Questionnaire (SMEQ)** [54] was administered. The SMEQ provides a post-task rating of the mental effort an individual expends in completing a task. It consists of a single numeric scale from 0 to 150, with nine labels from “Not at all hard to do” to “Tremendously hard to do.” Participants moved a slider on a computer to the point in the scale that represented their judgment of task difficulty. The software calculated the scale value selected by the participant. Higher values indicate greater perceived task difficulty.

4.1.4 Results. Results showed access to WB telephone audio improved speech recognition (in quiet) over speech recognition using typical NB telephone audio. In addition, WB audio resulted in lowering the perceived mental effort expended during completion of the speech recognition task compared to that expended for NB audio.

Speech Recognition: A paired t-test was performed between the conditions of NB and WB audio. WB audio resulted in better speech recognition compared to NB audio. The difference in percent words correct was statistically significant [$t(41) = 4.33, p = 9.364 \times 10^{-5}$]. The observed average of differences between WB audio and NB audio was 10.4%. Table 1 above shows mean results.

Mental Effort: A paired t-test was performed between the conditions of NB and WB audio. WB audio resulted in lower perceived mental effort compared to NB audio. The difference in

Table 2. Mean Subjective Mental Effort Statistics for WB x NB Audio (n = 42)

Codec	SMEQ	Std Dev	Std Error
AMR-WB 23.85	38.2	29.0	4.5
AMR-NB 12.20	54.5	32.4	5.0

Higher scores imply more effort.

perceived mental effort (SMEQ) was statistically significant [$t(41) = 3.63$, $p = 7.863 \times 10^{-4}$]. The observed average of differences between WB audio and NB audio was 16.3 points. Table 2 above shows mean results.

4.1.5 Discussion. Experiment 1 suggests that at optimum quality (i.e., the highest available data rates) under carefully controlled conditions (i.e., simulated phone audio over speakers at known levels in quiet), WB audio results in significantly better speech understanding, as well as significantly lower expended mental effort. This suggests that the AMR-WB codec adopted for implementing wireless mobile wideband audio could provide improved accessibility. However, mobile telephony, by its nature, does not take place under carefully controlled conditions. Acoustic noise, such as car noise, wind and the conversations of others, is often present in the environment during mobile device use, and the survey respondents identified environmental noise as a major barrier to telecommunications. For hearing individuals, wideband audio has been shown to increase speech quality and, in some cases, intelligibility in noise [17]. Considering the WB benefit shown for listening in quiet among people with hearing loss, the question arises whether increasing the telephone audio bandwidth could increase intelligibility in noise for individuals with hearing loss as well.

In videotelephony, the addition of a video signal to the mobile telephone speech signal in quiet is known to provide speech enhancement benefits, which are similar to the benefits of wideband audio in this experiment, if the two signals are synchronized [36]. However, mobile video telephony is also susceptible to acoustic noise in the environment. This raises the question as to whether wideband audio can provide additional benefits to videotelephony in quiet and whether it can help to mitigate any detrimental effects of environmental noise. These two questions provide the motivation for Experiment 2, which considers the impact of both video and noise.

4.2 Experiment 2: Codec Audio Bandwidth, Video and Noise

The follow-up research questions for Experiment 2 were: (1) Is the wideband vs narrowband audio effect as robust when listening alone in the presence of noise, and (2) To what degree is the effect present when visual speech information is combined with audio information as it would be in videotelephony in both quiet and noisy environments.

4.2.1 Participants. Testing was completed with a subgroup of 20 CI users from Experiment 1. Of the 20 individuals with hearing loss, 13 were women and 7 were men, with an average age of 57.3 years (ranging from 31–86 years). All participants had at least two to three years of self-reported hearing device use. All individuals used their cochlear implants during testing, with 11 bilateral CI users, 4 unilateral CI users and 5 individuals who used a CI in one ear and an HA in the other ear. Self-reported hearing loss ranged from severe to profound across both ears. Most participants (17) reported profound hearing loss in the ear(s) with a CI.

4.2.2 Materials. As in Experiment 1, stimuli for Experiment 2 were drawn from CASPER. Because the CASPER sentences each contain speech in audio and visual (lipreading) formats, we were able to use the same stimuli from the same corpus for evaluating NB and WB speech during audio-only (A-only) and audiovisual (A-V) testing. The audio and video signals in the CASPER sentence sets are synchronized, and the video is encoded with the Cinepak codec at a resolution

Table 3. Experimental Conditions of Presentation Mode, Environment and Audio Bandwidth

Condition	Presentation Mode	Environment	Audio Bandwidth
1	A-only	Quiet	NB
2	A-only	Quiet	WB
3	A-only	Noise	NB
4	A-only	Noise	WB
5	A-V	Quiet	NB
6	A-V	Quiet	WB
7	A-V	Noise	NB
8	A-V	Noise	WB

of 360×240 at 30fps with a bit rate of 7 Mbit/s. Sixteen sentence sets were used to prepare the stimuli, each spoken by the female speaker, with two different sets used for each test condition. The noise included a mix of speech babble (multiple voices) and environmental sounds typical of a large open meeting room where people can gather at tables.

Stimulus preparation: The eight conditions tested with associated presentation mode, environment and audio bandwidth used for this experiment are shown in Table 3 above. For presentation mode, the stimuli were presented either audio-only or audio-plus-video. The video in the audio-visual conditions were transcoded from Cinepak to the H.264 codec at a bit rate of 768 kbit/s for 30 fps and was perfectly synchronized with the audio signal during playback. Transcoding was required, as the Cinepak codec was not supported by the software we used to control the stimulus and noise playback in this round of experiments.

Note that some set-ups and screens induce drift between the audio and the video, so the encoding of the videos may need to adjust the audio delay relative to the video in order to achieve perfect synchronization during playback. In this specific experiment, which used Windows 7 for playing the stimuli, playing back recorded video of a clapper and examining the visual time of contact relative to the audio peak for the clap revealed that audio showed up 100 ms earlier than what was encoded in the video. We delayed the audio by 100 ms during the encoding to compensate for this effect and achieve perfect synchrony during playback. Different hardware and operating system configurations would require a different adjustment during encoding [36].

The listening environment was either quiet or with noise present. For the conditions in which noise was present, the relative presentation level of the speech signal compared to the level of the noise in the environment was calibrated to be +10 dB (see also the next section on method for details). The audio bandwidth used either AMR-NB or AMR-WB coding at the maximum bit rates as in Experiment 1.

4.2.3 Method. The test set up simulated cell phone listening with video at a size typical of 3.5-inch cell phone screens. The audio-video stimuli were presented on a 19" flat panel that had a cardboard mask over laid with simulated wireless phone controls and a rectangle cut out for presenting the video at the same dimensions as a typical smaller iPhone screen (Figure 9). Audio output of the speech signal was via external speakers, placed directly in front of the participant at 0° azimuth and, analogous to Experiment 1, sufficiently close to each other so that the sound appeared to come from a single source. The participants were seated comfortably, at a viewing distance of 12" from the bridge of the nose to the screen. All stimuli were played back through QuickTime, with the window placed such that the video matched the position of the cutout on the screen. The experimenter viewed the playback controls on a mirrored second screen. In audio-only testing conditions, the screen stayed black.

Two speakers played the noise (speech babble plus environmental noise), with each channel starting at a different point in the noise recording during playback. These uncorrelated noise

Fig. 9. Experimental setup for audio-visual testing showing the speaker setup and the screen with the cardboard mask. The mask contains a simulated iPhone, with the stimulus video showing through a cutout. Not drawn to scale.

Fig. 10. Experimental setup for testing in noise showing the speaker setup for noise presentation relative to the test set up for audio-visual testing.

sources provided a better surround sound experience than would correlated noise. The speakers were placed at $\pm 60^\circ$ off of 0° azimuth and approximately two-thirds of the way to both side walls from the participant (Figure 10). Both the noise and speech were calibrated at the approximate location of the subject's head. The frequent peaks of the speech were set at ~ 65 dB SPL, and the noise at 55 dB SPL in order to create the +10 dB S/N ratio (i.e., a reasonably noisy environment).

As in Experiment 1, participants repeated the sentences that they heard while using their hearing device(s). The participants' responses were scored as the number of words correctly repeated per set (out of 102 total words) for each condition. Presentation of conditions was counterbalanced across subjects. As in the previous experiment, following the completion of testing for one sentence set/one condition, the SMEQ was administered.

4.2.4 Results. The results showed an overall advantage for WB audio compared to NB audio in terms of speech recognition and mental effort. The addition of synchronized, visual speech information to the audio also significantly improved speech recognition and reduced mental effort. The WB audio advantage was greater for the audio-only condition than the audio-visual condition. Noise, on the other hand, both reduced speech recognition and increased mental effort regardless of whether participants were only listening or also had access to lipreading while listening. The negative impact of the noise was greater, however, when listening alone. And while WB audio lessened the negative impact of noise when listening alone, speech recognition was still poorer than for either NB or WB audio in quiet.

Speech Recognition: A repeated measures, three-way ANOVA for words correct showed significant main effects of the factors audio bandwidth (NB vs WB, $F(1,19) = 10.7$, $p < 0.004$), presentation mode (audio vs audio+video, $F(1,19) = 50.1$, $p < 0.000$), and environment (quiet vs noise, $F(1,19) = 73.8$, $p < 0.000$), with significant two-way interactions between presentation mode and environment ($F(1,19) = 25.5$, $p < 0.000$) and between presentation mode and audio bandwidth ($F(1,19) = 11.2$, $p < 0.003$). No other interactions were significant. For the interaction between presentation

Fig. 11. Mean number of words recognized and standard error for NB and WB audio bandwidth, grouped by listening environment and presentation mode ($n = 20$).

mode and environment, noise had a significantly greater negative impact on speech understanding regardless of audio bandwidth when participants were listening alone compared to when listening was accompanied by high quality, synchronized visual information. Noise reduced average speech understanding by approximately 25 percentage points for the audio-only condition, while for the audiovisual condition, mean speech understanding reduced on the order of only about 10 percentage points and remained above 80% words correct. In noise, average speech understanding when listening alone did not reach even 50% words correct. WB audio significantly improved average speech understanding over NB audio when listening-only compared to the audio-visual presentation mode regardless of the environmental conditions. Figure 11 above shows mean results.

Mental Effort: A repeated measures, three-way ANOVA for SMEQ ratings also showed significant main effects of the factors presentation mode ($F(1,19) = 53.6, p < 0.000$), environment ($F(1,19) = 41.6, p < 0.000$), and audio bandwidth ($F(1,19) = 6.8, p < 0.018$). The only significant interaction was between presentation mode and environment ($F(1,19) = 11.3, p < 0.003$). As with speech recognition, the noisy environment had a significantly greater negative impact on mental effort when listening alone than when listening was accompanied by visual information regardless of audio bandwidth. Even so, expenditures of mental effort were greater for participants listening alone in quiet than listening in noise when visual speech information was available. Concerning audio bandwidth, WB audio overall, without regard to presentation mode or environment, required lower expenditures of mental effort on the speech recognition task than did NB audio. Figure 12 below shows mean results.

4.2.5 Discussion. It is clear from both the speech recognition and mental effort data why survey respondents identified environmental noise as a major barrier to telecommunications. Speech recognition for listening alone was reduced by approximately 25 percentage points in noise, and the speech recognition task increased from ‘rather hard to do’ to ‘very hard to do.’ While the noisy environment also reduced performance in the videotelephony condition, its impact was considerably less compared to the quiet environment both in terms of speech recognition and expenditures

Fig. 12. Mean SMEQ ratings and standard error for NB and WB audio bandwidth, grouped by listening environment and presentation mode ($n = 20$).

of mental effort, which also shows up in the interaction effect between presentation mode and environment in the ANOVA. Still, the noisy environment affects the audio signal, presumably decreasing the availability and robustness of the auditory cues for integration with the visual speech cues, thereby increasing the demands of the recognition task.

The wideband audio advantage observed in Experiment 1 with optimum audio quality (i.e., the highest available data rates) was also present in this experiment. However, the advantage was greater in the audio-only condition compared to the audio-visual condition in terms of speech recognition performance, perhaps because of the already very robust lipreading enhancement effect provided by the addition of a synchronized visual signal. For ratings of subjective mental effort, the wideband audio advantage was present regardless of presentation mode or the presence of environmental noise. This suggests that although task performance did not suffer as much with the presence of noise when a bimodal input was provided, the noise still exerted a cost in terms of expenditures of mental effort that wideband audio was able to ameliorate to some degree.

Experiment 1 suggests that at optimum quality (i.e., the highest available data rates) under carefully controlled conditions (i.e., simulated phone audio over speakers at known levels in quiet), WB audio results in significantly better speech understanding, as well as significantly lower expended mental effort. Experiment 2, however, showed that the addition of WB audio shows diminishing returns when video is added and may not provide a sufficient advantage in noise to be the sole solution for better accessibility, but rather needs to work in concert with noise reduction techniques. Therefore, the remainder of the experiments were focused on the delivery of audio alone in quiet to identify further aspects of mobile audio quality that impact accessibility. In Experiment 3 we explore data bandwidth, which has been shown to impact audio quality for hearing individuals [1, 19]. Experiment 4 explores the network impairment of packet loss that can cause audio quality to suffer.

4.3 Experiment 3: Codec Audio Bandwidth and Bit Rate

The follow-up research question for Experiment 3 was: (1) Can the effect of WB audio be replicated not just at the highest data rates (i.e., AMR-WB 23.85 and AMR-NB 12.20), but also at lower data

Table 4. Experimental Conditions of Audio Codec and Bit Rate

Condition	Codec	Bandwidth	Bit Rate (kbit/s)
1	G.711 u-law	NB	64.00
2	AMR-NB	NB	5.90
3	AMR-NB	NB	12.20
4	AMR-WB	WB	12.65
5	AMR-WB	WB	23.85
6	AMR-WB	WB LP (lowpass filtered)	23.85

rates (i.e., AMR-WB 12.65 and AMR-NB 5.90) more representative of mobile networks, and (2) How does the quality of mobile audio codecs compare to that found on the PSTN (i.e., G.711 u-law). Finally, a low-pass version of the AMR-WB high bit rate codec was tested. This condition examined whether any change in performance with the AMR-WB codec at 23.65 kbps, compared to the others, was due to the higher bit rate or extended audio bandwidth used for encoding.

Additionally, the careful control of the sound levels was relaxed by switching from a simulated phone environment to a real wireless phone, and letting participants self-select their **most comfortable listening level (MCL)** on the phone. We also expanded the participant pool from CI users to a mix of CI and HA users. As mentioned previously, all participants were required to pass a listening test to ensure the audibility of tones with frequencies 4 and 5 kHz, and thereby verify access to the WB audio frequencies (see also the Section 4 overview).

4.3.1 Participants. Testing was completed with a group of 36 cochlear implant and hearing aid users. Of the 36 individuals with hearing loss, 23 were women and 13 were men, with an average age of 48 years (ranging from 18–73 years). All participants had at least one year of self-reported hearing device use. Fourteen individuals used their CIs during testing, while the other 22 used their HAs. 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 three reported moderate hearing loss, one reported mild hearing loss and one reported not knowing their degree of hearing loss. All CI users reported profound hearing loss.

4.3.2 Materials. As in Experiments 1 and 2, stimuli for Experiment 3 were drawn from CASPER. However, six different sentence sets, each spoken by two speakers (one male and one female), were used to prepare the stimuli, with two different sets used for each test condition (i.e., there were 12 sets of files across two speakers total). Each participant received one of the two speakers across all conditions. A separate set of CASPER sentences were used to train participants on the procedure and to establish an individual's MCL for telephone listening.

Stimulus preparation: The conditions with associated codecs and bit rates used for this experiment are shown in Table 4 above. G.711 u-law, a NB codec, provides baseline toll quality speech (i.e., quality comparable to a long-distance call placed over the PSTN). The next four conditions used AMR-NB and AMR-WB audio at the maximum bit rates for the highest possible quality (as in Experiment 1), and bit rates more typical of wireless mobile networks at the time of testing. The sixth condition was a **low-pass (LP)** filtered version of AMR-WB audio at the highest quality (i.e., 23.65 kbps) with a cut off frequency of 3500 Hz. The 3500 Hz cutoff was realized through an 8th-order Butterworth filter, in order to produce a high rate of roll off and a narrow transition band while having a maximally flat filter.

All sentence sets were processed in a similar way as in Experiment 1 for each of the six test conditions (cf. Experiment 1: Codec Audio Bandwidth), except for an additional preprocessing

Fig. 13. Experimental setup with iPhone 4S. The crossbar is holding the phone in the position that the participant selected for listening during the MCL setting. The Velcro headband holds the head immobile relative to the crossbar.

step. Because we intended the experiment to be representative of a real-world wireless phone to wireless phone call, it was necessary to ensure that the stimuli reflect the typical frequency response of a phone's microphone, as if a conversation partner had spoken into a real phone. To this end, a send mask, which is a type of equalizer with the filter curve described in the 3GPP TS 26.131 specification for testing phone audio [2], was applied to each sample prior to transcoding the original audio from PCM to AMR-NB and AMR-WB (as in Experiments 1 and 2 via ffmpeg and the OpenCore-AMR libraries). The sentence sets for each condition were counterbalanced across subjects.

4.3.3 Method. As mentioned above, in this experiment we switched from a simulated phone to a real phone held up at the participants' ear. 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 the researchers was used to control presentation of all stimuli in the correct order, and to control the phone volume settings via an external keyboard. The custom app ensured that the playback of the stimuli happened through the phone's ear speaker, rather than the speakerphone/media speakers typically found at the bottom of phone screens.

To minimize the risk of **radiofrequency (RF)** interference with the hearing device, no mobile or WiFi network connections were active on the phone during testing. Only Bluetooth was on for keyboard control. The phone was placed in a normal use position at the microphone of a participant's hearing device. To further eliminate RF interference, the phone screen was turned off by the custom app when it detected placement near the participant's ear via the phone's built-in proximity sensor. An adjustable stand was used to position and hold the handset. The stand was used in order to reduce fatigue, maximize coupling and maintain consistent handset positioning, thereby providing a constant input level to a hearing device's microphone during the course of testing. 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 hearing device's microphone and the phone's speaker for best-case acoustic coupling, as shown in Figure 13.

A Bluetooth keyboard paired with the phone was used by the testers to interact with the custom playback app on 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 as described in the following procedure and locked on the phone for the remainder of testing. While participants held the phone to their ear or the 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 whether they wanted the tester to increase or decrease the VC 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 locked. Note that the volume of the participants' hearing devices remained unchanged throughout the entire session – in fact, some hearing devices no longer offer manual volume controls.

As in the previous experiments, participants listened to and then repeated each sentence that they heard, and researchers scored their responses for the number of words correctly repeated in each sentence. Following presentation of all 12 sentences for a given condition, the **Mean Opinion Score (MOS)** was administered. 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. 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. As in the previous experiments, presentation of conditions was counter-balanced across subjects.

4.3.4 Results. A one-way ANOVA with repeated measures was performed separately for the speech recognition ($F(5,175) = 8.93, p < 10^{-6}$) and speech quality (MOS) ($F(5,175) = 17.3, p < 10^{-6}$) data, indicating in both cases that the means for the six audio codec conditions were not all equal. Further significance testing was done on the nine *a priori* pairwise comparisons planned from our research questions using the Tukey's HSD test, which computes a standardized Q score and controls for multiple comparisons.

It should be noted that the speech recognition data were highly positively skewed due to very high overall performance levels. Between 5 to 15 participants for each audio codec condition had maximum or near maximum speech recognition scores of 98–100 % correct. The high proportion of participants having maximal scores suggests a ceiling effect [13] occurred for speech recognition, especially for the WB audio codec conditions. While the arcsine transform for percent data was applied, and the degree of skewness was reduced; the data remained skewed. Therefore, while we report the results for speech recognition in Figures 14 and 15 and Table 5, they should be interpreted with caution, and speech quality, via the MOS scores, should be used to draw conclusions.

Overall, results showed that WB audio again provided benefits over NB audio for both speech recognition and speech quality. For the same coding strategy (AMR), WB audio @ 12.65 and 23.85 kbps produced better speech quality and speech recognition than NB audio @ 5.90 and 12.20 kbps, respectively. Additionally, higher bit rates produced better speech quality for both NB and WB audio than their lower bit rate counterparts. For AMR-NB audio, a higher bit rate also produced better speech recognition.

The much higher bit rate for G.711 NB audio (64 kbps) produced neither better speech quality nor speech recognition than AMR-NB @ 12.20 kbps, although G.711 did show better speech recognition and quality than AMR-NB @ 5.90 kbps. G.711, like the other NB codecs (AMR-5.9 and AMR-12.2), produced poorer speech quality and speech recognition than the two AMR WB codecs (AMR-12.65 and AMR-23.85).

Low-pass filtering the higher bit rate WB audio (23.85 kbps) reduced speech quality, but not speech recognition compared to unfiltered WB audio of the same bit rate.

Fig. 14. Mean percent words correctly recognized and standard error across participants for narrowband and wideband audio. G.711 is used on the PSTN. The other conditions are used on mobile telephony and grouped by the bit rates, low bit rates and the best-quality bit rates allowed by the respective codecs. AMR-WB 23.85 LP is the lowpass-filtered wideband audio encoding used to mimic narrowband audio at the best-quality bit rate supported by AMR-WB.

Fig. 15. Mean speech quality (MOS) and standard error across participants for narrowband and wideband audio. G.711 is used on the PSTN. The other conditions are used on mobile telephony and grouped by the bit rates, low bit rates and the best-quality bit rates allowed by the respective codecs. AMR-WB 23.85 LP is the lowpass-filtered wideband audio encoding used to mimic narrowband audio at the best-quality bit rate supported by AMR-WB.

4.3.5 Discussion. Experiment 3 showed that the WB audio advantage over NB still holds with bit rates typical of mobile telephony, with a real phone as opposed to simulated, and when participants are allowed to position the phone and determine MCL themselves. However, we can only fully interpret the codec differences on the basis of perceived speech quality because the analysis of speech recognition data was compromised due to the observed ceiling effect, particularly for the WB codecs. Codec differences with regard to perceived speech quality also showed that higher bit rates afford improvements, regardless of audio bandwidth.

The quality of the NB higher bit rate audio codec, AMR NB-12.20, used on mobile networks was essentially equivalent to the NB audio codec, G.711, used on the PSTN, while its lower bit

Table 5. Pairwise Comparisons ($n = 36$) with Post Hoc Significance Testing for Speech Recognition and MOS, for Conditions Displayed in Figure 14 and Figure 15

Comparisons	% Words Correct df(35)	MOS df(35)
1 NB 5.95 $\times$ NB 12.2	$t = 3.03, p < 0.0046$	$t = 2.25, p < 0.0309$
2 WB 12.65 $\times$ WB 23.85	$t = 0.77, p < 0.4474$ (n.s.)	$t = 2.94, p < 0.0057$
3 NB 5.95 $\times$ WB 12.65	$t = 5.00, p < 0.0001$	$t = 4.79, p < 0.0001$
4 NB 12.2 $\times$ WB 23.85	$t = 2.94, p < 0.0057$	$t = 6.81, p < 0.0001$
5 G.711 $\times$ NB 5.95	$t = 2.16, p < 0.0378$	$t = 2.58, p < 0.0142$
6 G.711 $\times$ NB 12.2	$t = 0.57, p < 0.5721$ (n.s.)	$t = 0.27, p < 0.7859$ (n.s.)
7 G.711 $\times$ WB 12.65	$t = 2.96, p < 0.0054$	$t = 4.07, p < 0.0003$
8 G.711 $\times$ WB 23.85	$t = 2.83, p < 0.0077$	$t = 5.61, p < 0.0001$
9 WB 23.85 $\times$ WB 23.85 LP	$t = 1.38, p < 0.1776$ (n.s.)	$t = 5.05, p < 0.0001$

(n.s. = not significant).

rate version, AMR NB-5.90, was poorer in quality. Given these results, the much higher bitrate, 64 kbits/s, and lossless nature of G.711 does not appear to have conferred any additional advantage over the lower bitrate, lossy codec, suggesting a limit to the speech quality improvements afforded by higher bit rate NB codecs for individuals with hearing loss.

The contribution of coding quality to the WB audio advantage was examined with the low-pass filtered WB-23.85 condition. In this condition, the extended frequency region above 3500 Hz was removed, but the higher bit rate encoding remained below 3500 Hz. The reduced quality of low-pass filtered WB AMR-23.65 compared to unfiltered WB AMR-23.65 supports the idea that bandwidth, rather than coding quality, is responsible for the observed WB audio advantage. Additionally, WB coding regardless of bitrate was superior to G.711. These results provide additional support of the importance of extended audio bandwidth for improved speech quality in telephony for individuals with hearing loss who have access to this frequency region.

To better understand the impact of the codec conditions examined in this experiment on speech recognition, this experiment should be replicated in the future using more challenging stimuli to reduce potential ceiling effects; we also did apply this lesson learned to Experiment 4.

4.4 Experiment 4: Codec Audio Bandwidth and Network Packet Loss

Experiment 4 expanded on the previous experiments to include a new audio quality condition involving the network impairment of packet loss, and at the same time replicated the previously examined audio quality factor of codec audio bandwidth. The research question for Experiment 4 was: (1) What is the impact of network packet loss on speech recognition, mental effort, and sound quality ratings for narrowband and wideband codecs. An additional question was: (2) How do these impacts differ between people with hearing loss and hearing people?

In the previous experiment, a ceiling effect was observed for the wideband audio stimuli for a number of participants. This suggests that the CASPER sentence sets were not challenging enough for these participants. Therefore, for this experiment, we employed the IEEE Harvard sentence set [26]. The sentences have low predictability and pose much greater challenges for speech recognition and are commonly used in voice telephony audio testing. We partnered with AT&T Labs to encode the Harvard stimuli, so as to ensure that the audio encoding exactly mirrored the behavior of real-world mobile handsets and networks, and that the packet loss mirrored conditions found in the real world.

4.4.1 Participants. Testing of NB vs WB telephone speech at various levels of packet loss was completed with a group of 36 cochlear implant and hearing aid users and 12 hearing individuals all

Table 6. Experimental Conditions of Audio Codec, Bit Rate, and Packet Loss, with a Full Factorial Design Across Bandwidth and Packet Loss

Condition	Codec	Bandwidth	Bit Rate (kbit/s)	Packet Loss
1	AMR-NB	NB	5.90	0%
2	AMR-NB	NB	5.90	3%
3	AMR-NB	NB	5.90	20%
4	AMR-WB	WB	12.65	0%
5	AMR-WB	WB	12.65	3%
6	AMR-WB	WB	12.65	20%

under the age of 50 years old. Of the 36 individuals with hearing loss, 25 were women and 11 were men, with an average age of 51 years (ranging from 22–79 years). All participants had at least two years of self-reported hearing device use. Twelve individuals used their CIs during testing, while the other 24 used their HAs. Self-reported hearing loss ranged from mild to profound across both ears. In the test ear, most HA users reported moderately-severe or severe hearing loss, while all CI users reported profound hearing loss.

4.4.2 Materials. Stimuli for this experiment were drawn from the IEEE Harvard sentence lists [26], which as mentioned above pose a much greater listening challenge than CASPER. These sentences are a collection of 72 lists of 10 sentences that are phonetically balanced, using specific phonemes at the same frequency they appear in English. Because these sentences 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, twelve sentence lists 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.

Recordings of all Harvard sentences were provided by the Sense Synergy company and included two male and two female speakers per sentence. All four speakers were fluent, native English speakers with typical speech and no discernable accents. 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 sentences were used to train participants on the procedure and to establish an individual's MCL for telephone listening.

Stimulus preparation: The conditions associated with codecs, bit rates and packet loss are shown in Table 6 above. Two baseline conditions and four conditions of reduced audio quality due to packet loss were prepared, for a full factorial design with audio bandwidth and packet loss as the factors. The baseline conditions included NB and WB audio with no packet loss at typical data rates used in mobile networks, determined in consultation with AT&T. The other four conditions involved two levels of bursty packet loss, 3% and 20%, for each baseline condition. The worst-case quality levels in a managed mobile network and over unmanaged Wi-Fi Internet calling, respectively, determined the percentages of packet loss selected. These conditions set upper/lower bounds on the effects of packet loss on speech quality likely to be experienced in mobile calling.

AT&T Labs carried out all signal processing for speech coding and injection of packet loss. 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. Each concatenated sentence set was processed using the NB and WB encoding strategies. Silence suppression (DTX, see the glossary in Appendix A) 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 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 codecs. Conceptually, 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 [19] with Gamma set at 0.8. This model utilizes a two-state Markov model approach with four degrees of freedom and is widely used to generate impairments that simulate transmission failures in real-time services over telecom networks. Within this model, using a gamma consolidates the four degrees of freedom into correlated ones; details on this are provided in Appendix B. Conceptually, 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 [29], 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 losslessly converted to PCM audio to support playback on an iPhone. 100 ms of silence was added to the beginning and end of each sentence to avoid any clipping induced by playback delays, and levels were equalized as in the previous experiments.

4.4.3 Method. As in Experiment 3, participants used their preferred ear, preferred phone positioning (in the case of hearing participants, to their ear) and self-selected speech MCL through adjusting the volume on the phone. The setup of the hardware and software was identical to that of Experiment 3, as well. The same custom app on an iPhone 4S was used and controlled through a Bluetooth keyboard. As before, the phone was held in position with a stand and a Velcro headband (cf. Experiment 3: Codec Audio Bandwidth and Bit Rate - Method; and Figure 13).

The procedure was analogous to that of the previous experiment (cf. Experiment 3: Codec Audio Bandwidth and Bit Rate). Prior to testing, all participants received training on the entire procedure. The speech MCL was established and locked in the phone's volume settings for the remainder of the testing. For each study participant, speech recognition was tested using one set of IEEE sentences for each of the six audio quality conditions via listening to and repeating sentences, providing ratings; followed by reconfirming audibility of band noises. Participants provided both SMEQ and MOS ratings.

Each administration of one condition took seven minutes. Presentation of conditions was counterbalanced across subjects, and the sentence sets chosen for each condition were also counterbalanced across subjects. Additionally, the presentation order of sentences was randomized, and each condition featured each of the two male and two female speakers in equal measure, where each participant encountered different permutations of speakers across different sentences. To guard against bias, a double-blind procedure was used in which neither researchers administering the experiment, nor the participants, were aware of which conditions were being evaluated for any given sentence set.

4.4.4 Results. Overall, results showed for people with hearing loss that WB audio continues to confer benefits over NB audio in the areas of speech recognition, perceived mental effort, and speech audio quality ratings. Hearing participants saw benefits, too, especially under packet loss conditions. Additionally, participants with hearing loss had poorer speech recognition, higher perceived mental effort and lower ratings of speech audio quality, compared to their hearing

Table 7. Results for Testing Audio Bandwidth and Packet Loss on Participants with Hearing loss (HL) and Hearing Participants (NH)

Cond.	HL % Words Correct	% Std Err	HL SMEQ	SMEQ Std Err	HL MOS	MOS Std Err	NH % Words Correct	% Std Err	NH SMEQ	SMEQ Std Err	NH MOS	MOS Std Err
NB 0%	77.6	3.5	50.3	6.4	3.4	0.2	98.1	0.8	14.3	3.1	3.8	0.2
NB 3%	73.6	3.3	55.8	5.4	3.1	0.2	95.3	1.1	26.2	3.9	3.3	0.2
NB 20%	42.9	3.1	83.8	5.8	2.0	0.2	83.0	2.6	60.6	7.1	2.2	0.3
WB 0%	81.2	3.4	37.1	5.2	3.9	0.2	99.2	0.9	2.3	0.9	4.7	0.1
WB 3%	77.3	3.5	47.6	5.7	3.4	0.2	98.5	0.9	14.8	3.0	3.9	0.2
WB 20%	51.2	3.8	84.1	5.3	2.0	0.2	91.5	1.7	45.2	5.3	2.6	0.3

Fig. 16. Mean percent words correct with respect to the packet loss and audio bandwidth factors, for both hearing (H) participants and participants with hearing loss (HL). There were significant main effects for packet loss and audio bandwidth for both H and HL, but only a significant interaction for H.

counterparts, regardless of audio codec bandwidth and degree of packet loss. There was no ceiling effect present for participants with hearing loss, in contrast to Experiment 3. The detailed results are shown in Table 7 above.

Speech Recognition: For participants with hearing loss, a repeated measures, two-way ANOVA for words correct showed significant 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 ($F(2,70) = 1.44, p < 0.243$). Pairwise comparisons, carried out separately for mean differences between the two levels of the main effect of audio bandwidth and the three levels of the main effect of packet loss, were all significant at the 0.05 level. No other pairwise comparisons were carried out because there were no significant interaction effects.

For the hearing participants, a repeated measures, two-way ANOVA for words correct showed significant 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 ($F(2,22) = 5.82, p < 0.009$); see Figure 16. Pairwise comparisons were significant at the 0.05 level for mean differences between NB and WB audio at the 3% and 20% levels of packet loss but not at 0% packet loss. For WB audio, pairwise comparisons were significant at the 0.05 level for mean differences between

Fig. 17. Mean SMEQ scores with respect to packet loss and audio bandwidth, for both hearing (H) participants and participants with hearing loss (HL). There were significant main effects for packet loss and audio bandwidth for both H and HL, but no significant interaction.

0% and 20% and 3% and 20% packet loss, but not between 0% and 3% packet loss. While for NB audio, all pairwise comparisons of 0%, 3% and 20% packet loss were significant at the 0.05 level.

Mental Effort: For participants with hearing loss, a repeated measures, two-way ANOVA for SMEQ ratings showed significant 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 ($F(2,70) = 1.75$, $p < 0.181$). Pairwise comparisons, carried out separately for mean differences between the two levels of the main effect of audio bandwidth and the three levels of the main effect of packet loss, were all significant at the 0.05 level. No other pairwise comparisons were carried out because there were no significant interaction effects.

For the hearing participants, a repeated measures, two-way ANOVA for SMEQ ratings showed significant 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 ($F(2,22) = 0.329$, $p < 0.723$); see Figure 17. Pairwise comparisons, carried out separately for mean differences between the two levels of the main effect of audio bandwidth and the three levels of the main effect of packet loss, were all significant at the 0.05 level. No other pairwise comparisons were carried out because there were no significant interaction effects.

Mean Opinion Score: For participants with hearing loss, a repeated measures, two-way ANOVA for MOS ratings showed significant 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 ($F(2,70) = 3.22$, $p < 0.046$). Pairwise comparisons were significant at the 0.05 level for mean differences between NB and WB audio at the 0% level of packet loss but not at 3% or 20% packet loss. For WB audio, pairwise comparisons were significant at the 0.05 level for mean differences between all levels of packet loss. While for NB audio, pairwise comparisons were significant at the 0.05 level for mean differences between 0% and 20% and 3% and 20% packet loss, but not between 0% and 3% packet loss.

For the hearing participants, a repeated measures, two-way ANOVA for MOS ratings showed significant 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 ($F(2,22) = 1.06$, $p < 0.363$); see Figure 18. Pairwise comparisons, carried out separately for mean differences between the two levels of the main effect of audio bandwidth and the three levels of the main effect of

Fig. 18. Main and interaction effects plot for MOS with respect to the packet loss and audio bandwidth factors, for both hearing (H) participants and participants with hearing loss (HL). There were significant main effects for packet loss and audio bandwidth for both H and HL, but only a significant interaction for HL.

packet loss, were all significant at the 0.05 level. No other pairwise comparisons were carried out because there were no significant interaction effects.

4.4.5 Discussion. Overall, participants with hearing loss had poorer speech understanding, higher expenditures of mental effort, and lower perceived speech quality 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 higher levels of packet loss. Otherwise, speech understanding for this group was similarly high for narrowband and wideband audio when no packet loss was present. This phenomenon shows up in the interaction effect for hearing participants with respect to their recognition scores. 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 compared to 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 in the presence of packet loss, which did not occur for the hearing participants. This shows up in the interaction effect for MOS among participants with hearing loss. Although a similar trend shows in the data for the subjective mental effort, it did not meet the thresholds for significance.

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 the first three experiments have been replicated in this experiment, despite the employment of much more challenging stimuli.

5 DISCUSSION

5.1 Main Findings

We have shown for 68 cochlear implant and 46 hearing aid users, for a total of 114 participants with hearing loss, that wideband audio confers significant benefits in the areas of speech recognition, reduced perceived mental effort, and perceived audio quality. These findings have been replicated across four studies. The codec and bit rate typical of wireless mobile telephony – AMR-NB 5.90 – results in both poorer speech recognition and poorer perceived audio quality than higher-quality narrowband encodings prevalent on VoIP and PSTN. For wideband audio, the typical AMR-WB12.65 supports recognition similar to higher quality bit rates, but perceived audio quality is still better with higher bit rates. Packet loss and noise result in significantly degraded performance; however, the advantage of wideband over narrowband audio still holds, especially in terms of packet loss on managed networks that limit packet loss to 3%. Finally, we have also confirmed that among hearing people, wideband audio confers benefits, especially under degraded network conditions.

5.2 Applications

Overall, these studies strongly suggest that ubiquitous wideband audio in telecommunications is a win-win for hearing people and people with hearing loss alike – speech quality and accessibility are both significantly improved over narrowband audio. AMR-WB has been deployed by carriers in the US in their 4G networks and handsets operating **Voice over Long-Term Evolution (VoLTE)**. The improvement in data rates accompanying the deployment of LTE provides the opportunity to increase the bandwidth for speech at very little additional cost in terms of data usage.

Implementers of telecommunications systems should favor wideband audio over narrowband audio, and technical/policy standards should make wideband support required (e.g., as is already done in Section 508 in the US, and ETSI EN301 549 in Europe), rather than recommended (e.g., the next-generation emergency calling specification in North America [48]). Telecommunication providers also should make sustained efforts to ensure interoperability of wideband audio calls across environments and carriers, which has languished in the US since a 2014 petition in front of the Federal Communications Commission [62]. In particular, efforts need to be undertaken to ensure that wideband audio works in calls across competing wireless carriers, and in calls from/to a wireless carrier and a fixed-line VoIP service provider.

If the narrowband audio codec AMR-NB must be supported in legacy telecom, bit rates higher than 5.90 kbit/s should be considered. Additionally, managed networks with quality-of-service guarantees, so as to minimize packet loss, can potentially offer better accessibility for voice calls than services that operate on the open internet.

5.3 Limitations

Study limitations are related to the stimuli and data collection methods used in the experiments. With regard to stimuli, we used a limited set of audio codecs. There are other codecs that are widely implemented in cellular and web-based telecommunications that should be explored (such as G.722, G.729 and Opus, which recently was explored with CI users [23]). The packet loss conditions used static percentages of packet loss. In real-world networks, packet loss rates are not static

but vary over time. Our video for the audiovisual condition was both high quality and synchronous with the audio. Real world network conditions can degrade video quality and affect synchronicity, where the benefit of lip-reading is reduced or lost entirely. Further testing using audiovisual stimuli should include common types and degrees of real-world network degradation that can impact the visual signal channel.

As for data collection methods, the stimuli used were designed for receptive listening tasks. While phone conversations between two people do require receptive listening skills, speech understanding is supported by context. This context includes not just what precedes and follows a spoken sentence, but the setting of the conversation and the various dimensions of the conversational partners' roles and each individual's background knowledge. Receptive listening tasks, while common types of performance indicators, may only be partially reflective of performance during phone conversations between two people. Additionally, our testing used microphone coupling (i.e., the audio goes from phone speaker to hearing device microphone) and a fixed handset position. While our survey results indicated that it was the most common coupling method used for telephone listening, it does not represent the only coupling method employed by hearing device users. We did not explore other coupling methods, such as telecoil or Bluetooth coupling. Likewise, we did not explore the variations that can occur in the acoustic signal received at the hearing device's microphone when the handset is held at the ear without the fixed positioning we used in the experiments. Addressing each of these limitations is important to understanding how our findings translate to real-world use.

5.4 Future Work

Future work also should consider what bit rates are appropriate for people with hearing loss for the Opus codec, which is widely implemented in web-based audio and video calling. Additionally, Opus has been adopted as a requirement in the 2017 Section 508 Refresh in the United States, which means that equipment procured under federal regulatory standards will support this codec in the future. Additional work is needed to validate up-and-coming AI-based audio codecs, such as Microsoft's Satin.

Now that we better understand the impact of speech coding and network impairments, appropriate follow-up studies should explore these technical parameters in more ecologically valid tasks and environments. Although this paper provides clear evidence toward improving receptive telephone listening, the relationship between listening and the ability of people with hearing loss to hold phone conversations is poorly understood, and merits further investigation using conversational-type tasks. In particular, the effect of these parameters should be explored in two-way conversations, where calling partners have the opportunity to detect and repair conversation breakdowns induced by poor audio quality and the characteristics of a person's hearing loss.

6 CONCLUSION

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. Likewise, using higher bit rates for wideband codecs and narrowband codecs may improve the accessibility of mobile telephony. Wideband audio may confer benefits when high-quality video is added to support the audio with lip-reading although its importance may not be on the same level as that for audio-only telecommunications. In degraded conditions, such as the presence of environmental noise or network packet loss, wideband audio also may confer benefits. However, in both these situations, we have seen diminishing returns. When the degradation becomes increasingly severe, as was the case with packet loss, the degree of benefit from WB audio may reduce. And, when the impact of

the degradation is significant, as was the case when listening alone in noise, absolute performance may remain low in spite of the improvements afforded by WB audio.

All these experiments involved receptive listening tasks, and their generalization to two-way real-world conversations still remains to be determined. The results, however, suggest that there are concrete actionable steps that implementers of telecommunications systems (both wireless mobile and web/internet-based) can take to improve accessibility and usability for all, including people with hearing loss.

APPENDICES

A GLOSSARY

- **AMR-NB** – Adaptive Multirate Narrowband, used to encode narrowband audio on cellular networks.
- **AMR-WB** – Adaptive Multirate Wideband, used to encode wideband audio on cellular networks.
- **CI** – shorthand for cochlear implant.
- **DTX** – shorthand for discontinuous transmission; a method of saving bandwidth during moments of silence.
- **Acoustic Feedback Control** – a means of reducing or eliminating the audible whistle created when the acoustic output of a hearing aid escapes the ear canal, reenters the hearing aid through its microphone, and is reamplified.
- **Frequent Peaks of Speech** – the points of highest amplitude in the acoustic waveform of a speech signal.
- **G.711 u-law** (Toll-Quality Speech) – uncompressed narrowband audio transmitted via internet-based telephony; comparable to best-case analog phone networks (PSTN).
- **HA** – shorthand for hearing aid.
- **HAC** – certain types and levels of built-in performance in wireline and wireless phones that is required by federal law in order for these phones to work with hearing aids.
- **MCL** (Most Comfortable Loudness Level) – The volume setting deemed most comfortable by a person.
- **Packet Loss Concealment** (PLC) – a method to smooth over packet loss; for example, by repeating and attenuating the audio information from the previous packet.
- **PCM** – Pulse Code Modulation, lossless encoding of audio.
- **Public Switched Telephone Network** (PSTN) – the analog wired phone and first- to third-generation cellular networks.
- **RF Interference** (Radio Frequency Interference) – interference caused in the hearing device by the electromagnetic emissions of digital wireless phones.
- **Streamer** – a separate, intermediary device that connects hearing devices to sound sources for communication and entertainment; the connection between the sound source and the streamer is via Bluetooth while the connection between the streamer and the hearing device is via telecoil coupling or a proprietary RF connection.
- **Telecoil** – a small coil of wire placed in a hearing device that acts as a receiver for magnetic signals. Frequently used by a person with hearing loss to reduce the impact of environmental noise while making a phone call.
- **Transcoding** – the process of converting audio or video from one codec to another, in order to improve compression or avoid codec incompatibilities in editing or playback software.
- **VoIP** (Voice over Internet Protocol) – internet-based telephony, as well as fourth-generation cellular telephony.

- **VoLTE** (Voice over Long-Term Evolution) – internet protocol-based transmission of voice over fourth-generation cellular networks.

B GILBERT-ELLIOT MODEL FOR PACKET LOSS

The telecommunications industry frequently relies on reference code provided by ITU-T [27]. This code provides software and a manual for simulating a variety of network and encoding conditions, including those of packet loss. The current version of this reference code includes samples for the **Gilbert Elliot (GEC)** model, which also has been incorporated into the Linux Netem module, by way of the token bucket filter.

The GEC has four degrees of freedom; however, it is also common to consolidate these into a single parameter called “gamma,” which is what AT&T provided for the experiments described in this paper. The actual parameters for the GEC can be calculated, following the exposition on pages 161–163 in the STLmanual.pdf file contained within the archive of the ITU-T sample code [27].

As per Equations 11.5 and 11.6 in this manual, the conversions between gamma and the four degrees of freedom are as follows:

1. P_G : fixed at 0
2. P_B : fixed at 0.5
3. $P = 2 \cdot (1 - \gamma) \cdot \text{PLR}$ where PLR is the packet loss rate
4. $Q = (1 - \gamma) \cdot (1 - 2 \cdot \text{PLR})$

In our experiments we used packet loss rates of 3% and 20%, and $\gamma = 0.8$, so the PLR values are, respectively, 0.03 and 0.2. Or specifically, for 3% packet loss:

1. $P = 2 \times (1 - 0.8) \times 0.03 = 2 \times 0.2 \times 0.03 = 0.012$
2. $Q = (1 - 0.8) \times (1 - 2 \times 0.03) = 0.2 \times 0.94 = 0.188$

And for 20% packet loss:

1. $P = 2 \times (1 - 0.8) \times 0.2 = 2 \times 0.2 \times 0.2 = 0.08$
2. $Q = (1 - 0.8) \times (1 - 2 \times 0.2) = 0.2 \times 0.6 = 0.12$

The corresponding command line syntax to configure Linux’s Netem module is

tc qdisc add dev <device> <priority band> netem loss gemodel P Q PB PG

1. For 3% packet loss this translates into: **tc qdisc add dev <device> <priority band> netem loss gemodel 1.2% 18.8% 50% 0%**
2. For 20% packet loss, this translates into: **tc qdisc add dev <device> <priority band> netem loss gemodel 8% 12% 50% 0%**

REFERENCES

- [1] 3GPP TR 26.935 V8.0.0 Technical Report. 2008. *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Packet-switched conversational multimedia applications; Performance characterisation of default codecs (Release 8)*. Retrieved May 1, 2019 from <https://www.3gpp.org/ftp/Specs/rel8/26/26935-800.pdf>.
- [2] 3GPP TS 26.131 V12.3.0 Technical Specification. 2014. *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Packet-switched conversational multimedia applications; Terminal acoustic characteristics for telephony; Requirements (Release 12)*. Retrieved February 26, 2021 from <https://www.3gpp.org/ftp/Specs/rel12/26/26131-c30.pdf>.
- [3] Doug L. Beck and Jes Olsen. 2008. Extended bandwidths in hearing aids. *Hearing Review* 15, 11 (2008), 22–26.
- [4] John G. Beerends, Christian Schmidmer, Jens Berger, Matthias Obermann, Raphael Ullmann, Joachim Pomy, and Michael Keyhl. 2013. Perceptual objective listening quality assessment (POLQA), the third generation ITU-T standard

- for end-to-end speech quality measurement part I—Temporal alignment. *Journal of the Audio Engineering Society* 61, 6 (2013), 366–384
- [5] Arthur Boothroyd. 1987. CASPER: Computer-Assisted speech perception evaluation and training. In *10th Annual Conference on Rehabilitation Technology* Washington, DC.
 - [6] Marc A. Brennan, Ryan McCreery, Judy Kopun, Brenda Hoover, Joshua Alexander, Dawna Lewis, and Patricia G. Stelmachowicz. 2014. Paired comparisons of nonlinear frequency compression, extended bandwidth, and restricted bandwidth hearing aid processing for children and adults with hearing loss. *Journal of the American Academy of Audiology* 25, 10 (2014), 983–998. <https://doi.org/10.3766/jaaa.25.10.7>
 - [7] James R. Bress, Steven Whitesell, and Tony Jasionowski. 2013. Telecommunications Industry Association (TIA): Developing standards for accessibility. Presented to the TR-41.3.14 Accessibility Working Group. Retrieved May 1, 2019 from http://standards.tiaonline.org/standards/_committees/documents/DevelopingStandardsForAccessibility.pdf.
 - [8] Jack Browne. 2018. Comparing narrowband and wideband channels. *Microwaves & RF*. Retrieved February 26, 2021 from <https://www.mwrf.com/technologies/systems/article/21848973/comparing-narrowband-and-wideband-channels>.
 - [9] Andrés Buzo, Augustine H. Gray, Robert M. Gray, and John Markel. 1980. Speech coding based upon vector quantization. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 5 (1980), 562–574. <https://doi.org/10.1109/TASSP.1980.1163445>
 - [10] Neva Cherniavsky, Anna Cavender, Richard Ladner, and Eve A. Riskin. 2007. Variable frame rate for low power mobile sign language communication. *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility* 163–170. ACM. <https://doi.org/10.1145/1296843.1296872>
 - [11] Teresa Y. C. Ching and Harvey Dillon. 2013. A brief overview of factors affecting speech intelligibility of people with hearing loss: Implications for amplification. *American Journal of Audiology* 22, 306–309. [https://doi.org/10.1044/1059-0889\(2013/12-0075\)](https://doi.org/10.1044/1059-0889(2013/12-0075))
 - [12] Jaehong Chon, Sam Whittle, Eve A. Riskin, and Richard Ladner. 2011. Improving compressed video sign language conversations in the presence of data loss. *2011 Data Compression Conference*, 383–392. IEEE. <https://doi.org/10.1109/DCC.2011.45>
 - [13] Azadeh Ebrahimi-Madiseh, Robert H. Eikelboom, Dona M. P. Jayakody, and Marcus D. Atlas. 2016. Speech perception scores in cochlear implant recipients: An analysis of ceiling effects in the CUNY sentence test (Quiet) in post-lingually deafened cochlear implant recipients. *Cochlear Implants International* 17, 2 (2016), 75–80. <https://doi.org/10.1080/14670100.2015.1114220>
 - [14] Bernd Edler, Andreas Buechner, Waldo Nogueira, and Frank Klefenz. 2020. Cochlear implant, device for generating a control signal for a cochlear implant, device for generating a combination signal and combination signal and corresponding methods. U.S. Patent 10,744,322, issued August 18, 2020.
 - [15] Tiago H. Falk, Vijay Parsa, Joao F. Santos, Kathryn Arehart, Oldooz Hazrati, Rainer Huber, James M. Kates, and Susan Scollie. 2015. Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools. *IEEE Signal Processing Magazine* 32, 2 (2015), 114–124. <https://doi.org/10.1109/MSP.2014.2358871>
 - [16] Federal Communications Commission. 2018. 2018–2019 TRS Rate Order. FCC #DA-18-680. Retrieved May 1, 2019 from <https://www.fcc.gov/document/2018-2019-trs-rate-order>.
 - [17] Hans W. Gierlich. 2005. Wideband speech communication-The quality parameters as perceived by the user. In *Proceedings of Forum Acusticum* Vol. 2005.
 - [18] Jérémie Guignard, Pascal Senn, Roger Koller, Marco Caversaccio, Martin Kompis, and Georgios Mantokoudis. 2019. Mobile internet telephony improves speech intelligibility and quality for cochlear implant recipients. *Otol Neurotol* 40, 3 (2019), e206–e214. <https://doi.org/10.1097/MAO.0000000000002132>
 - [19] Gerhard Haßlinger and Oliver Hohlfeld. 2008. The Gilbert-Elliott model for packet loss in real time services on the Internet. In *14th GI/ITG Conference-Measurement, Modelling and Evaluation of Computer and Communication Systems. 1-15 VDE*.
 - [20] Andrew Hines, Jan Skoglund, Anil Kokaram, and Naomi Harte. 2013. Robustness of speech quality metrics to background noise and network degradations: Comparing VisQOL, PESQ and POLQA. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3697–3701. <https://doi.org/10.1109/ICASSP.2013.6638348>
 - [21] Andrew Hines, Jan Skoglund, Anil C. Kokaram, and Naomi Harte. 2015. VisQOL: An objective speech quality model. *EURASIP Journal on Audio, Speech, and Music Processing* 2015, 1 (2015), 1–18. <https://doi.org/10.1186/s13636-015-0054-9>
 - [22] Reemt Hinrichs, Tom Gajecki, Jörn Ostermann, and Waldo Nogueira. 2019. Coding of electrical stimulation patterns for binaural sound coding strategies for cochlear implants. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2019*, 4168–4172. <https://doi.org/10.1109/EMBC.2019.8857271>

- [23] Reemt Hinrichs, Tom Gajecki, Jörn Ostermann, and Waldo Nogueira. 2021. A subjective and objective evaluation of a codec for the electrical stimulation patterns of cochlear implants. *The Journal of the Acoustical Society of America* 149, 2 (2012), 1324–1337. <https://doi.org/10.1121/10.0003571>
- [24] Christian Hoene, Berthold Rathke, and Adam Wolisz. 2003. On the importance of a VoIP packet. In *ISCA Tutorial and Research Workshop on Auditory Quality of Systems* 55–62.
- [25] Yi Hu, Qudsia Tahmina, Christina Runge, and David R. Friedland. 2013. The perception of telephone-processed speech by combined electric and acoustic stimulation. *Trends in Amplification* 17, 3 (2013), 189–196. <https://doi.org/10.1177/1084713813512901>
- [26] Institute of Electrical and Electronics Engineers. 1969. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics* 17, 3 (1969), 225–246.
- [27] International Telecommunications Union (ITU-T). 2019. G.191: Software tools for speech and audio coding standardization. Retrieved July 31, 2020 from <https://www.itu.int/rec/T-REC-G.191-201901-I/en>.
- [28] International Telecommunications Union (ITU-T). 2001. P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs Retrieved February 26, 2021 from <https://www.itu.int/rec/T-REC-P.862-200102-I/en>.
- [29] Wenyu Jiang and Henning Schulzrinne. 2000. Modeling of packet loss and delay and their effect on real-time multimedia service quality. In *Proc. NOSSDAV*.
- [30] Stephen Julstrom, Linda Kozma-Spytek, and Scott Isabelle. 2011. Telecoil-mode hearing aid compatibility performance requirements for wireless and cordless handsets: Magnetic signal-to-noise. *Journal of the American Academy of Audiology* 22, 8 (2011), 528–541. <https://doi.org/10.3766/jaaa.22.8.5>
- [31] Stephen Julstrom and Linda Kozma-Spytek. 2014. Subjective assessment of cochlear implant users' signal-to-noise ratio requirements for different levels of wireless device usability. *Journal of the American Academy of Audiology* 25, 10 (2014), 952–968. <https://doi.org/10.3766/jaaa.25.10.4>
- [32] Anna Chi Shan Kam, John Ka Keung Sung, Tan Lee, Terence Ka Cheong Wong, and Andrew van Hasselt. 2017. Improving mobile phone speech recognition by personalized amplification: Application in people with normal hearing and mild-to-moderate hearing loss. *Ear and Hearing* 38, 2 (2017), e85–e92. <http://dx.doi.org/10.1097/AUD.0000000000000371>
- [33] Laura J. Kepler, Mark Terry, and Richard H. Sweetman. 1992. Telephone usage in the hearing-impaired population. *Ear and Hearing* 13, 311–319.
- [34] Mead Killion and H. Gustav Mueller. 2010. Twenty years later: A new count-the-dots-method. *The Hearing Journal* 63, 1 (2010), 10–17.
- [35] Linda Kozma-Spytek. 2003. Hearing aid compatible telephones: History and current status. *Seminars in Hearing* 24, 1 (2003), 17–28.
- [36] Linda Kozma-Spytek, Paula Tucker, and Christian Vogler. 2013. Audio-visual speech understanding in simulated telephony applications by individuals with hearing loss. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* No. 6. ACM. <https://doi.org/10.1145/2513383.2517032>
- [37] Linda Kozma-Spytek, Paula Tucker, and Christian Vogler. 2019. Voice telephony for individuals with hearing loss: The effects of audio bandwidth, bit rate and packet loss. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 3–15. <https://doi.org/10.1145/3308561.3353796>
- [38] Charles J. Limb and Alexis T. Roy. 2014. Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research* 308, 13–26. <https://doi.org/10.1016/j.heares.2013.04.009>
- [39] Frank R. Lin, John K. Niparko, and Luigi Ferrucci. 2011. Hearing loss prevalence in the United States. *Arch Intern Medicine* 171, 20 (2011) 1851–1853. <https://doi.org/10.1001/archinternmed.2011.506>
- [40] Frank R. Lin, Roland Thorpe, Sandra Gordon-Salant, and Luigi Ferrucci. 2011. Hearing loss prevalence and risk factors among older adults in the United States. *J Gerontol A Biol Sci Med Sci*. 66A, 5 (2011), 582–590. <https://doi.org/10.1093/gerona/gle002>
- [41] Chuping Liu, Quen-Je Fu, and Shrikanth S. Narayanan. 2009. Effect of bandwidth extension to telephone speech recognition in cochlear implant users. *The Journal of the Acoustical Society of America* 125, 2 (2009), EL77–EL83. <https://doi.org/10.1121/1.3062145>
- [42] Carol L. Mackersie, Yingyong Qi, Arthur Boothroyd, and Nicole Conrad. 2009. Evaluation of cellular phone technology with digital hearing aid features: effects of encoding and individualized amplification. *Journal of the American Academy of Audiology* 20, 2 (2009), 109–118. <https://doi.org/10.3766/jaaa.20.2.4>
- [43] Georgios Mantokoudis, Claudia Dähler, Patrick Dubach, Martin Kompis, Marco D. Caversaccio, and Pascal Senn. 2013. Internet video telephony allows speech reading by deaf individuals and improves speech perception by cochlear implant users. *PLoS One* 8, 1 (2013), e54770. <https://doi.org/10.1371/journal.pone.0054770>
- [44] Georgios Mantokoudis, Patrick Dubach, Flurin Pfiffner, Martin Kompis, Marco Caversaccio, and Pascal Senn. 2012. Speech perception benefits of internet versus conventional telephony for hearing-impaired individuals. *Journal of Medical Internet Research* 14, 4 (2012), e102. <https://doi.org/10.2196/jmir.1818>

- [45] Georgios Mantokoudis, Martin Kompis, Patrick Dubach, Marco Caversaccio, and Pascal Senn. 2010. How internet telephony could improve communication for hearing-impaired individuals. *Otol Neurotol* 31, 7 (2010), 1014–1021. <https://doi.org/10.1097/MAO.0b013e3181ec1d46>
- [46] Steven C. Marcum, Erin M. Picou, and Thomas Steffens. 2017. Avoiding disconnection: An evaluation of telephone options for cochlear implant users. *International Journal of Audiology* 56, 3 (2017), 186–193. <https://doi.org/10.1080/14992027.2016.1247502>
- [47] National Academies of Sciences, Engineering, and Medicine (NASEM). 2016. Hearing health care for adults: Priorities for improving access and affordability. Retrieved April 27, 2019 from <http://nationalacademies.org/hmd/reports/2016/Hearing-Health-Care-for-Adults.aspx>.
- [48] National Emergency Number Association. 2016. Detailed Functional and Interface Standards for the NENA i3 Solution. NENA-STA-010.2-2016. Retrieved May 1, 2019 from https://www.nena.org/page/i3_Stage3.
- [49] Waldo Nogueira, Johannes Abel, and Tim Fingscheidt. 2019. Artificial speech bandwidth extension improves telephone speech intelligibility and quality in cochlear implant users. *The Journal of the Acoustical Society of America* 145, 3 (2019), 1640–1649. <https://doi.org/10.1121/1.5094347>
- [50] OpenCore-AMR. 2019. Library of OpenCORE framework implementation of adaptive multi rate narrowband and wideband (AMR-NB and AMR-WB) speech codec. Retrieved April 27, 2019 from <https://sourceforge.net/projects/opencore-amr/>.
- [51] Pew Research Center. 2019. Mobile fact sheet. Retrieved February 20, 2021 from <https://www.pewresearch.org/internet/fact-sheet/mobile/>.
- [52] Erin M. Picou and Todd A. Ricketts. 2013. Efficacy of hearing-aid based telephone strategies for listeners with moderate-to-severe hearing loss. *Journal of the American Academy of Audiology* 24, 1 (2013), 59–70. <https://doi.org/10.3766/jaaa.24.1.7>
- [53] Christina M. Roup, Gayla L. Poling, Kimberly Harhager, Ashok Krishnamurthy, and Lawrence L. Feth. 2011. Evaluation of a telephone speech-enhancement algorithm among older adults with hearing loss. *Journal of Speech, Language, and Hearing Research* 54, 5 (2011), 1477–1483. [https://doi.org/10.1044/1092-4388\(2011/10-0181\)](https://doi.org/10.1044/1092-4388(2011/10-0181))
- [54] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of three one-question, post-task usability questionnaires. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 1599–1608*. ACM. <https://doi.org/10.1145/1518701.1518946>
- [55] Astrid Schmidt-Nielsen and Karen R. Stern. 1985. Identification of known voices as a function of familiarity and narrow-band coding. *Journal of the Acoustical Society of America* 77, 2 (1985), 658–663. <https://doi.org/10.1121/1.391884>
- [56] Angeline Seeto and Grant D. Searchfield. 2018. Investigation of extended bandwidth hearing aid amplification on speech intelligibility and sound quality in adults with mild-to-moderate hearing loss. *Journal of the American Academy of Audiology* 29, 3 (2018), 243–254. <https://doi.org/10.3766/jaaa.16180>
- [57] Tomohiko Taniguchi, Shigeyuki Unagami, and Robert M. Gray. 1988. Multimode coding: A novel approach to narrow- and medium-band coding. *Journal of the Acoustical Society of America* 84, S12. <https://doi.org/10.1121/1.2025766>
- [58] Tomohiko Taniguchi, Yoshinori Tanaka, and Robert M. Gray. 1991. *Speech Coding with Dynamic bit Allocation (multimode coding)*. *Advances in Speech Coding 157–166*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-3266-8_16
- [59] Jessica J. Tran, Ben Flowers, Eve A. Risken, Richard Ladner, and Jacob O. Wobbrock. 2014. Analyzing the intelligibility of real-time mobile sign language video transmitted below recommended standards. *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* 177–184. <https://doi.org/10.1145/2661334.2661358>
- [60] Nancy Tye-Murray, Jacqueline Spry, and Elizabeth Mauze. 2009. Professionals with hearing loss: Maintaining that competitive edge. *Ear and Hearing* 30, 4 (2009), 475–484. <https://doi.org/10.1097/aud.0b013e3181a61f16>
- [61] Vanessa Vas, Michael A. Akeroyd, and Deborah A. Hall. 2017. A data-driven synthesis of research evidence for domains of hearing loss, as reported by adults with hearing loss and their communication partners. *Trends in Hearing* 21 (2017), 1–25. <https://doi.org/10.1177/2331216517734088>
- [62] Vishu R. Viswanathan, John Makhoul, Richard Schwartz, and A. W. F. Huggins. 1982. Variable frame rate transmission: A review of methodology and application to narrow-band LPC speech coding. *IEEE Transactions on Communications* 30, 4 (1982), 674–686. <https://doi.org/10.1109/TCOM.1982.1095523>
- [63] Voice Communication Exchange Committee (VCXC). 2014. Petition for Rulemaking. Filed with the FCC, GN Docket 13–5. February 25, 2014. Retrieved May 1, 2019 from <https://www.fcc.gov/ecfs/filing/6017604148>.
- [64] Jace Wolfe, Mila Morais Duke, Erin Schafer, George Cire, Christine Menapace, and Lori O'Neill. 2016. Evaluation of a wireless audio streaming accessory to improve mobile telephone performance of cochlear implant users. *International Journal of Audiology* 55, 2 (2016), 75–82. <https://doi.org/10.3109/14992027.2015.1095359>

Received July 2020; revised July 2021; accepted August 2021