Received May 10, 2017, accepted May 26, 2017, date of publication June 5, 2017, date of current version June 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2711489

A Smartphone-Based Multi-Functional Hearing Assistive System to Facilitate Speech Recognition in the Classroom

ALAN CHERN[1], YING-HUI LAI[2], (Member, IEEE), YI-PING CHANG[3,4], YU TSAO[5], (Member, IEEE), RONALD Y. CHANG[5], (Member, IEEE), AND HSIU-WEN CHANG[4]

[1]School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332 USA
[2]Department of Electrical Engineering, Yuan Ze University, Taoyuan 320, Taiwan
[3]Speech and Hearing Science Research Institute, Children's Hearing Foundation, Taipei 114, Taiwan
[4]Department of Audiology and Speech-Language Pathology, Mackay Medical College, New Taipei 252, Taiwan
[5]Research Center for Information Technology Innovation, Academia Sinica, Taipei 115, Taiwan

Corresponding author: Hsiu-Wen Chang (iamaudie@hotmail.com)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 103-2218-E-715-002, Grant MOST 104-2218-E-715-002, and Grant MOST 105-2218-E-155-014-MY2.

ABSTRACT In this paper, we propose a smartphone-based hearing assistive system (termed SmartHear) to facilitate speech recognition for various target users, who could benefit from enhanced listening clarity in the classroom. The SmartHear system consists of transmitter and receiver devices (e.g., smartphone and Bluetooth headset) for voice transmission, and an Android mobile application that controls and connects the different devices via Bluetooth or WiFi technology. The wireless transmission of voice signals between devices overcomes the reverberation and ambient noise effects in the classroom. The main functionalities of SmartHear include: 1) configurable transmitter/receiver assignment, to allow flexible designation of transmitter/receiver roles; 2) advanced noise-reduction techniques; 3) audio recording; and 4) voice-to-text conversion, to give students visual text aid. All the functions are implemented as a mobile application with an easy-to-navigate user interface. Experiments show the effectiveness of the noise-reduction schemes at low signal-to-noise ratios in terms of standard speech perception and quality indices, and show the effectiveness of SmartHear in maintaining voice-to-text conversion accuracy regardless of the distance between the speaker and listener. Future applications of SmartHear are also discussed.

INDEX TERMS Wireless assistive technologies, smartphone technologies, hearing assistive systems, speech recognition, classroom, e-health, m-health.

I. INTRODUCTION

Good acoustic characteristics are prerequisites to an effective and less stressful learning experience in the classroom. Information is predominantly exchanged via verbal communication in a typical classroom and therefore effective speech recognition through listening is crucial. It has been reported that listening, as opposed to reading, speaking, and writing, is the most occurred communication behavior among college students, who spend 55.4% of their total communication time on listening in the classroom [1].

The effectiveness of speech recognition through listening can be measured by the listening effort, defined as the attention and cognitive resources required for an individual to perform auditory tasks [2], [3]. It was reported that listening effort depends on age [3], [4] (i.e., older adults expend more listening effort than young adults in recognizing speech in noise) as well as on hearing conditions [5] (i.e., hearing-impaired children expend more listening effort than their normal-hearing peers in classroom learning). Furthermore, it was reported that considerable listening effort is required when listening at typical classroom signal-to-noise ratios (SNRs) [6]. Increased listening effort leads to increased listening and mental fatigue [5], [7].

Speech recognition is a process that occurs not only in the auditory modality but also visual modality. The role of visual information is particularly prominent when the perceived SNR of the auditory speech is less favorable. Visual cues can compensate for the limitations of auditory abilities of hearing-impaired individuals by enhancing speech recognition in noise and can reduce listening effort [8]. Speech converted to text, as a form of visual speech cues, could be beneficial for high school and college students with hearing impairment in obtaining lecture information [9]. Subtitling the video could improve comprehension of the contents and benefit hearing-impaired and normal-hearing individuals alike [10]. To increase students' access to educational resources and accommodate their educational needs, captioning, known as real-time speech-to-text service [11], is also provided in class. With modern technology advances, speech-to-text services have migrated to a computer-based delivery mode where service providers produce text on a computer as the teacher speaks to an automatic speech recognition (ASR) machine [11].

Factors that affect an individual's listening and speech recognition abilities in the classroom include the surrounding noise level, the reverberation time, and the acoustic quality of the transmitted speech. The American National Standard Institute (ANSI), in collaboration with the Acoustical Society of America, has specified an SNR of 15 dB or higher at the student's ear, a reverberation time of less than 0.6 seconds, and a maximum noise level of 35 dBA in an unoccupied classroom for an amiable classroom learning environment [12].

Frequency modulation (FM) systems [13] have been a routine clinical recommendation to improve speech recognition in noise. FM systems include two integral parts, namely, a transmitter and a receiver. The transmitter is placed near the speaker, and the receiver is placed near the user. The speaker's speech is picked up by the transmitter's microphone and transmitted to the receiver via FM signals. The benefits of FM systems on improving SNR and poor room acoustics have been proven in many studies [14]–[17].

With the advances of smartphone technologies, there is a potential to implement the working principles of FM systems on smartphones using wireless technologies for audio streaming. Mobile Based Assistive Listening System (MoBALS) [18] and Jacoti Lola Classroom [19], for example, provide smartphone-based assistive listening solutions that promise better affordability and availability and less stigma than the commercial FM systems. MoBALS and Jacoti Lola Classroom can be connected to users' hearing devices via direct audio input or used alone.

In this paper, we introduce the proposed smartphone-based multi-functional hearing assistive system (SmartHear) to facilitate speech recognition in the classroom. SmartHear requires only the user's smartphone, Bluetooth or WiFi connectivity, and our mobile application, and may potentially conform to the ANSI standards for classroom acoustics. SmartHear distinguishes itself from conventional FM systems and other existing smartphone-based implementations of FM systems as a novel new hearing assistive system of its kind. SmartHear not only transmits the voice of the speaker directly to the listener's ear to overcome reverberation and ambient noise effects, but also applies advanced noise-reduction techniques to improve the SNR of the received speech, supports audio recording to facilitate future audio reference or playback, and implements voice-to-text conversion to ease listening effort with visual aid. All these functions are implemented as a mobile application with an easy-to-navigate user interface. The target users of SmartHear include students with hearing loss, students with attention deficit hyperactivity disorder (ADHD), students in the foreign language classroom, or anybody who could potentially benefit from listening with better clarity in a classroom setting [20]–[22].

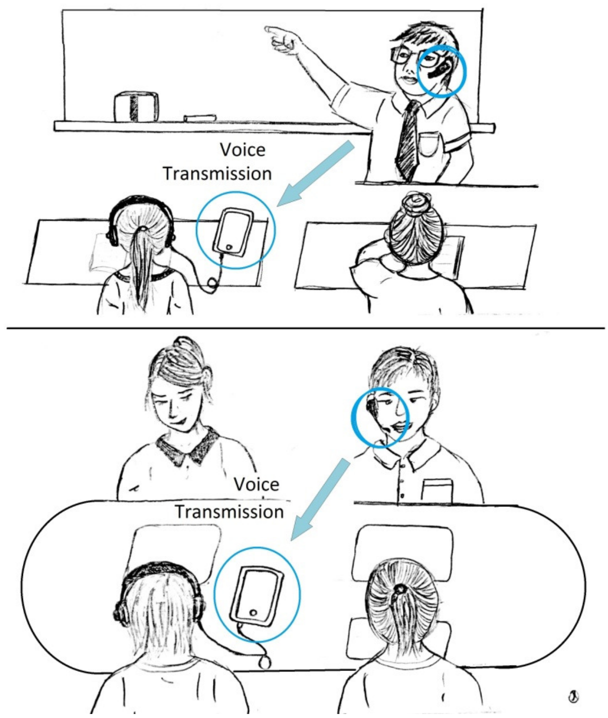


FIGURE 1. Exemplary use cases for SmartHear, including big classroom lecture (top) and small group discussion (bottom) scenarios. The teacher or speaker is equipped with a transmitter (here, a Bluetooth microphone) while the student listens through a paired receiver (here, a smartphone with the headphones).

The outline of this paper is as follows. Section II introduces the SmartHear system architecture. Section III describes the user interface and usage for SmartHear. Section IV presents the evaluation methods. Section V presents the experimental results and discussion. Finally, Section VI concludes the paper.

II. SmartHear SYSTEM CONCEPT, FUNCTIONS, AND IMPLEMENTATIONS

In this section, we describe SmartHear system concept, functions, and implementations. Exemplary application scenarios for SmartHear are portrayed in Fig. 1, including big classroom lecture and small group discussion scenarios. SmartHear operates in one of the two transmission modes: i) one-to-many voice transmission, by connecting a transmitting smartphone and multiple receiving smartphones via the WiFi multicast functionality, and ii) one-to-one voice transmission (as shown in Fig. 1), by connecting a smartphone and its coupled Bluetooth device via the Bluetooth technology. The voice of the teacher or speaker is picked up by the transmitter and then transmitted directly to the receiver(s) for listening. The use of wireless radio frequencies to transmit audio signals overcomes the reverberation and ambient noise effects and thereby enhances listening experience.

To specifically target the needs of students in a classroom setting and improve their learning experiences through enhanced listening, SmartHear incorporates the following features:

1) *Configurable transmitter/receiver assignment:* For the one-to-many transmission mode, any smartphone may be configured as the transmitter while the other smartphones in the same WiFi network as the receivers. For the one-to-one transmission mode, either the smartphone or the Bluetooth device may be configured as the transmitter while the other as the receiver (Fig. 1 illustrates the case where the Bluetooth microphone is configured as the transmitter).
2) *Advanced noise-reduction techniques:* The advanced noise-reduction algorithms computed on the smartphone mitigate the surrounding noise that is often present in the classroom and improve the perceived audio quality.
3) *Audio recording:* This function allows students to record the lectures or discussions and play back the audio later when needed.
4) *Voice-to-text conversion:* This function gives students visual text aid, which is particularly useful during small group discussions.

The implementation of these features is described as follows.

1) CONFIGURABLE TRANSMITTER/RECEIVER ASSIGNMENT

The one-to-many transmission mode requires a proper multicast configuration of the connecting WiFi network to function properly. During a transmission session, the voice of the speaker is first temporarily recorded by the smartphone, which converts the sound from analog to digital. Next, the smartphone converts the digital signal to radio waves for transmission. Through the Internet Group Management Protocol (IGMP), the smartphone and network router establish multicast group membership, allowing the smartphone to transmit the signal over a 2.4 GHz industrial, scientific, and medical (ISM) radio band to the router and then ultimately to the other smartphones registered as receivers in the same WiFi network. Finally, the receiving smartphones convert the radio signals back to analog to be played out for the listeners. The flow of the application programming interfaces (APIs) used to achieve such transmission on a smartphone is shown in Fig. 2. The Java/Android APIs AudioRecord, AudioTrack, DatagramPacket, and MulticastSocket are used. AudioRecord records audio captured by a designated input and stores the data temporarily. DatagramPacket and MulticastSocket are responsible for configuring the datagram settings and sending/receiving

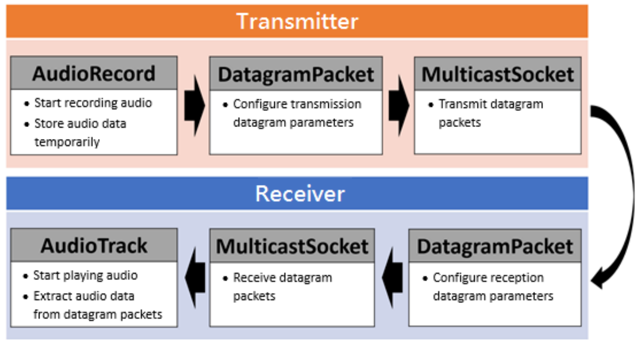


FIGURE 2. Java/Android API flow from transmitter to receiver device for the SmartHear one-to-many transmission mode.

TABLE 1. Technical configurations for SmartHear one-to-many transmission mode (all six rows) and one-to-one transmission mode (bottom four rows).

Parameter	Value
Input port	2010
Multicast address	239.255.1.1
Sampling rate	11.025 kHz
Audio channel	Mono
Audio format	PCM in 16 bits
Audio buffersize	640 bytes

datagram packets, respectively. Once the receiver obtains the data, AudioTrack extracts and plays the audio at a designated output. SmartHear implements configurable transmitter/receiver assignment with these APIs.

The technical configurations for SmartHear one-to-many transmission mode are summarized in Table 1. The network input port can be chosen from the range of 1024 to 49151, and 2010 is the selected port number for our application. The available multicast addresses range from 239.255.0.0 to 239.255.255.255, and SmartHear uses address 239.255.1.1. In terms of the audio specifications, there are four sampling frequency options, namely, 11.025, 16, 22.05, and 44.1 kHz. Since higher sampling frequency may result in longer latency, SmartHear uses 11.025 kHz sampling frequency. For the sound format, mono audio channel (as opposed to stereo) is used since it is more widely supported by Bluetooth devices. The audio format is the 16-bit pulse-code-modulated (PCM) digital signal, which is also supported by more audio devices as compared to 8-bit PCM. Finally, the audio buffersize is set to 640 bytes to avoid overwriting any buffered data.

The one-to-one transmission mode operates independently of the surrounding network infrastructure. Due to its self-containedness, it is the de facto setting for SmartHear. The transmission channel between the smartphone and the Bluetooth device is established again using the Android APIs AudioRecord and AudioTrack, with an additional AudioManager API, as shown in Fig. 3. Specifically,

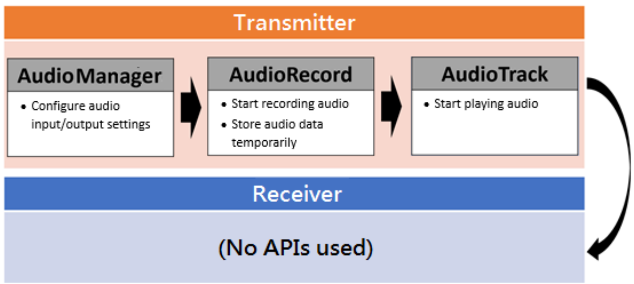


FIGURE 3. Android API flow from transmitter to receiver device for the SmartHear one-to-one transmission mode.

TABLE 2. Audio channel routing for the SmartHear one-to-one transmission mode.

	With Wired Earphones		Without Wired Earphones	
	Audio In	Audio Out	Audio In	Audio Out
Talk Mode	Default	Voice Recognition	Default	Voice Recognition
Listen Mode	Voice Call	Music	Music	Alarm

AudioRecord is coupled with AudioTrack to receive and then play out audio, while AudioManager is utilized to specify and switch between the hardware used for sound input and output (e.g., smartphone and Bluetooth device) to ultimately route the direction of audio flow as desired. Note that unlike the one-to-many transmission mode, the receiver here is a Bluetooth device rather than another smartphone, and therefore no APIs are needed on the reception end. The configuration settings in the bottom four rows of Table 1 also apply to the one-to-one transmission mode.

Generally, an Android device internally decides which hardware component is used for the reception/output of various audio types. For example, voice is usually played out from the smartphone built-in speaker, but when a Bluetooth or wired headset is connected, the sound is automatically routed to this headset instead. As a result, without correctly manipulating the audio channel, sound that is input into a headset microphone may get output by the headset itself rather than from the smartphone. Table 2 shows the audio channels selected for various setups. ''Talk Mode'' and ''Listen Mode'' here refer to configuring the smartphone as the transmitter and receiver, respectively. For the case of ''Talk Mode,'' the input audio is routed to ''Default,'' meaning that the smartphone uses the default microphone (i.e., the smartphone built-in microphone) for voice reception. The output audio from the Bluetooth device is routed to the ''Voice Recognition'' channel to override the default sound output location. Otherwise, when wired earphones are plugged into the smartphone, they preempt the Bluetooth device for outputting audio and sound might undesirably travel from the microphone to the wired earphones instead of the Bluetooth device. For the case of ''Listen Mode,'' similar logic applies. However, here, having wired earphones plugged in or not alters the routing configurations. Specifically, when wired earphones are plugged into the smartphone, the input audio into the Bluetooth device is routed to the ''Voice Call'' channel and the output audio from the smartphone is routed to the ''Music'' channel. When wired earphones are not plugged into the smartphone, the input audio into the Bluetooth device is routed to the ''Music'' channel and the output audio from the smartphone is routed to the ''Alarm'' channel.

For the one-to-one transmission mode, the two different transmitter/receiver assignments present different advantages and suit different needs. Speaking into the smartphone and listening through a Bluetooth headset is suitable when only the teacher possesses a smartphone while the student does not. Since the SmartHear application is readily available on the Android application store, the student only needs to prepare a Bluetooth headset in order to start a listening session. This option is especially useful for younger students who do not have a smartphone. On the other hand, if the student has a smartphone, having the teacher speak into a Bluetooth microphone and the student listen through the smartphone presents several advantages as the student now maintains control of the smartphone. First, since the background noise level in a classroom can greatly fluctuate, the student can easily manage the audio volume at will. Second, the student can start and pause the application at any desired time (e.g., the application can be paused when recess starts). Third, in case of an emergency, the student does not have to interrupt the class to retrieve the smartphone from the teacher.

The transmission ranges for one-to-many and one-to-one transmission modes are determined by the supporting wireless technologies. Most Bluetooth devices support a range of about 10 m, while some can provide a longer range (up to 100 m). This transmission range is sufficient for most practical applications considered in this paper (e.g., in the classroom). WiFi typically supports a longer range (10–100 m), which can be used when listening in a large auditorium or similar venues where a WiFi network is available with a proper multicast configuration.

2) ADVANCED NOISE-REDUCTION TECHNIQUES

SmartHear's noise-reduction functionality advances listening clarity and promotes student use of the application. The classroom environment can range from being very quiet to very noisy, and this function is particularly indispensable for noisy situations. The primary goal of noise reduction is to remove the noise components of noise-corrupted speech signals and accordingly enhance the speech quality and perceptual quality [23], [24]. Conventional noise reduction methods involve two integral parts: noise tracking and gain (filter) estimation, which respectively aim to calculate the statistics of noise and speech and compute a gain function. The estimated gain function is then applied to the noise-corrupted speech signals to restore the clean speech signals. More recently, data-driven-based noise reduction algorithms, such as deep learning algorithms [25]–[27], have been proposed

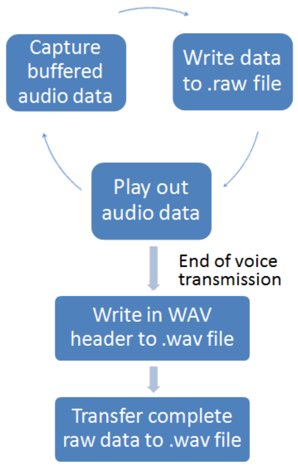


FIGURE 4. **Flowchart of implementing the audio recording function in SmartHear.**

and proven effective in challenging conditions. While deep learning-based models may achieve superior performance, conventional speech entrenchment methods are more computationally feasible for implementation on mobile devices. Specifically, we employ generalized maximum *a posteriori* spectral amplitude (GMAPA) algorithm [28] in our application. GMAPA was shown advantageous, objectively and subjectively, over other related noise-reduction algorithms (e.g., maximum likelihood spectral amplitude (MLSA) and MAPA) [28]. The implementation of noise-reduction on smartphones requires a Fast Fourier Transform (FFT) library to accomplish the necessary signal processing tasks. After the buffered audio data are received on the smartphone, the signal is transformed to the frequency domain so the noise-reduction algorithm and calculations can be carried out. When the noise-reduction function is activated, this step is repeated continuously for each buffered audio data chunk that is input into the smartphone.

3) AUDIO RECORDING

This function allows the audio data to be saved on the smartphone for future playback or reference, which is useful for students whose primary language is not the language used in the classroom teaching or discussion. As depicted in Fig. 4, a raw file is first created in the internal memory to keep record of all the audio data. Then, upon finishing a voice transmission session, a WAV audio file is generated by adding 44 bytes of WAV header data to the beginning of the raw data. The WAV header contains some specifications regarding the audio, such as the sampling rate, recording resolution (bits per sample), and audio format [29]. The audio recording function also includes a safeguard feature that creates backup files every 30 minutes in addition to the complete audio file. This prevents loss of the entire recording data in unforeseen circumstances such as smartphone crashing or battery dying. As previously mentioned, SmartHear transmits audio at a sampling rate of 11.025 kHz with 16-bit PCM

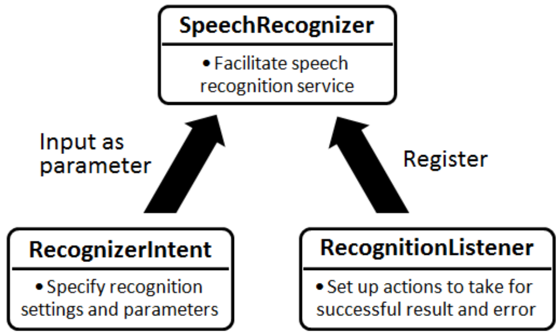


FIGURE 5. **Three main APIs used for implementing the voice-to-text function in SmartHear.**

format. Such settings achieve good tradeoff between the speech quality and the required memory space for recording. The file size of the resulting WAV audio recording is approximately 2.5 megabytes (MB) per minute, which translates to about 150 MB for a one-hour lecture. Thus, along with the backup files, which occupy roughly the same memory space, a complete recording of a typical one-hour lecture requires about 300 MB memory space. As modern smartphones commonly provide built-in or external memory in the order of tens to hundreds of gigabytes (GB), a long continuous recording or several shorter recordings can easily be managed.

4) VOICE-TO-TEXT CONVERSION

The SmartHear's voice-to-text function utilizes three ready-to-use Android APIs provided by Google for ASR, namely, SpeechRecognizer, RecognizerIntent, and RecognitionListener. As shown in Fig. 5, the SpeechRecognizer class is used to initialize and obtain a reference to a SpeechRecognizer object, which is used to facilitate the recognition service. Then, the SpeechRecognizer registers the RecognitionListener which sets up the actions to take after the SpeechRecognizer receives a response. The two most relevant responses are when the recognizer receives a successful result and when an error has occurred, in which cases the implementation of SmartHear simply displays the resultant text and the error message, respectively. Finally, the SpeechRecognizer initiates the reception of speech by passing in an instance of RecognizerIntent, which specifies speech recognition settings and parameters including language model and recognition language. The language model selected is the free form model because it is suited for recognition of sentences rather than individual word terms. Currently, the voice-to-text function can only support sentence-by-sentence transcription due to the limitations of the SpeechRecognizer API [30]. Thus, the current sentence-based voice-to-text function may be more suitable for small group discussions or one-on-one sessions where the speaker can talk at a slower and more predictable pace. Besides, these Android APIs only allow reception/processing of audio on a single smartphone, and therefore the voice-to-text function is currently available via

FIGURE 6. **SmartHear's user interfaces related to the "Broadcast" function. The initial screen prompts the user to choose between "One to Many" and "One to One." After choosing "One to Many" (respectively, "One to One"), the second screen prompts the user to connect to WiFi (respectively, Bluetooth). Once connected, the third screen is displayed, which consists of various control configurations for the voice transmission session. The fourth screen shows the text box dialog which is prompted at the end of a recording session.**

Bluetooth connection with a Bluetooth device but not via WiFi connection with other smartphones.

III. SmartHear USER INTERFACE AND USAGE

In this section, we introduce SmartHear mobile application's user interface that allows easy navigation and operation of the aforementioned functions.

The SmartHear mobile application user interface consists of a title bar at the top and a selection menu at the bottom, with the space in between being the area for function controls and information display. The title bar shows the application name followed by the current functionality chosen. The bottom menu offers three options to select from, namely, "Broadcast," "Recordings," and "Voice to Text," which are interfaces for features 1 & 2, feature 3, and feature 4 described in Section II, respectively. These three options are each displayed with an icon in addition to the corresponding text. The icon-text pairs are colored gray by default and turn to an orange color when selected in order to allow the user to easily identify which option is active. The center area of the screen displays different interfaces based on the selected function, and contains controls and instructions to guide the user. The displayed language of the application is based on the language setting of the smartphone. Currently, the available languages are English and Traditional Chinese; any other language setting defaults the display to English.

A. BROADCAST

When the SmartHear application starts up, the "Broadcast" function is selected as the default. Fig. 6 shows all sequence of screens that are associated with "Broadcast." The first screen prompts the user to choose between "One to Many" and "One to One," which represent one-to-many and one-to-one transmission modes, respectively. Depending on which connection mode was selected, the user is led to a connection screen to connect to either WiFi or Bluetooth. If the device already has the necessary connection, the application automatically moves on to the control screen. On the control screen, the user is presented with five control options: "Talk," "Listen," "Record," "Reduce Noise," and "START." The "Talk" and "Listen" options allow users to switch between transmitting audio through their smartphones ("Talk") and receiving audio through their smartphones ("Listen"). The "Record" and "Reduce Noise" options configure whether the audio will be recorded and whether noise-reduction processing will be applied, respectively. While either "Talk" or "Listen" may be selected at any given time, the "Record" and "Reduce Noise" options are not mutually exclusive, i.e., both can be toggled on or off at the same time. Similar to

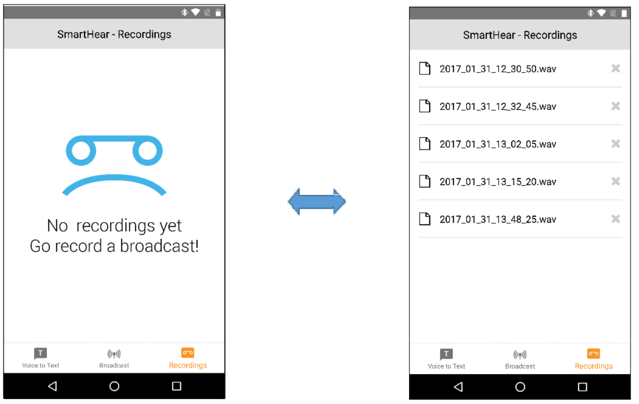


FIGURE 7. SmartHear's user interfaces related to the "Recordings" function. If there are no recordings available, the interface shows a message that informs the user (left). If there exists at least one audio recording file, the interface shows a list of previously saved audio files which can be either played or deleted (right).

the bottom selection menu icons, these four options turn orange when selected and are gray otherwise. When the green "START" button is pressed, a voice transmission session begins according to the selected options, and the button becomes a red "STOP" button (not shown in Fig. 6). During the active transmission session, the user can freely switch between "Talk" and "Listen," as well as toggle "Record" and "Reduce Noise" on/off any time. When ready to end the session, the user presses the red "STOP" button, and the button then reverts back to the green "START" button again. If the "Record" option had been toggled on at least once during the transmission session, the user is prompted a text box dialog to enter a file name to save the audio file. The file name is defaulted to "yyyy_MM_dd_HH_mm_ss.wav," which represents year, month, date, hour (in the 24-hour format), minute, and second information of the beginning of the recording.

B. RECORDINGS

Fig. 7 shows the screens for the "Recordings" function. Depending on whether the user has saved any voice recording files, this screen either indicates that there are no recordings available or shows a list of saved recording files. Clicking on a file name plays the audio file, while tapping the "X" symbol next to each file deletes the file.

C. VOICE TO TEXT

Fig. 8 shows the screens for the "Voice to Text" function. Since the usage of this function requires a Bluetooth connection as is the case for "Broadcast" with Bluetooth, when "Voice to Text" is selected, the user is prompted to connect to a Bluetooth device if a connection is not established yet. The control screen shows a green "START" button and a space reserved for displaying the result text after a voice-to-text session. When the "START" button is pressed, the application processes the next segment of speech received through the paired Bluetooth microphone and displays the corresponding text on the smartphone screen. The top-right menu button contains "SAVE," "CLEAR," and "Browse Saved Files" options. The "SAVE" option outputs all the results into an accessible text file for later use. Similar to the audio recording naming scheme, this option also prompts a dialog box for user input, where the default file name is again "yyyy_MM_dd_HH_mm_ss.txt." The "Browse Saved Files" option displays a list of saved text files (not shown in Fig. 8), where the user can click on the file name to open the file or select the "X" symbol to delete the file. The "CLEAR" option clears out all result text currently displayed on the screen.

The SmartHear mobile application is available for download on Google Play, and a video introduction to SmartHear is available on YouTube.

IV. EVALUATION METHODS

A. SmartHear's VOICE TRANSMISSION

We focus on the one-to-one transmission mode for performance evaluation. The performance of SmartHear's voice transmission is evaluated based on two standard evaluation indices, namely, Hearing-Aid Speech Perception Index (HASPI) and Hearing-Aid Speech Quality Index (HASQI) [31], [32]. These two indices are objective measures designed for accurate predictions of speech intelligibility and quality for individuals with or without hearing impairment listening to noisy speech signals. The score ranges of HASPI and HASQI are from 0 to 1, with a higher score indicating better speech intelligibility and quality. Both these indices are auditory/cochlear model-based and account for changes due to hearing loss.

Our experimental setup, as shown in Fig. 9, involves a PC connected to a speaker playing consecutive English utterances from the IEEE corpus [24], a Jabra Clipper Bluetooth headset (with an incorporated microphone) [33] stationed 5 cm in front of the speaker for voice reception, an HTC One (M8) smartphone [34] running the SmartHear application with a set of Apple EarPod earphones plugged in, and a Kemar [35] (wearing the earphones) placed 2.7 m away from the speaker. Inside the Kemar's ear, a microphone is installed to record the sound from the earphones. The PC-speaker system is calibrated to a fixed volume such that the released audio is set to a 75-dB sound pressure level (SPL). The entire process is done inside a rectangular IAC model 120A-1 standard double-walled audiometric test room of approximate dimensions $6 \times 5.5 \times 2.25$ m^3 to prevent any outside noise. This overall environment is intended to mimic a classroom setting, where the speaker represents the teacher who speaks into the Bluetooth microphone, while the Kemar represents a student who sits in the front row in the classroom listening via the smartphone.

The utterances are transmitted via SmartHear application's "Broadcast" function, with the "Reduce Noise" option either turned on or off (hereafter termed the "noise reduction mode" and "standard mode," respectively). The two

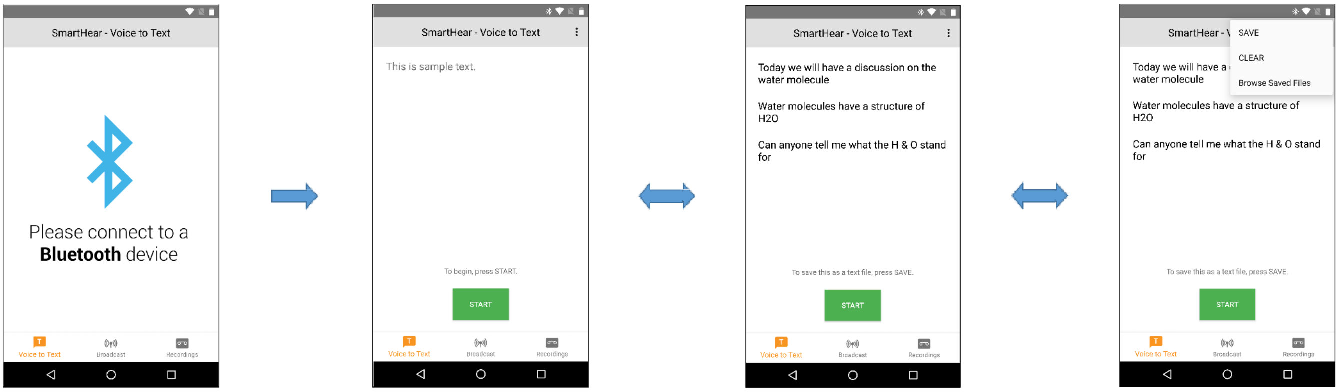


FIGURE 8. **SmartHear's user interfaces related to the "Voice to Text" function. The initial screen prompts the user to connect to a Bluetooth device. Once connected, the second screen shows up, which contains the controls for this function. The third screen contains the speech recognition result texts being displayed. The last screen shows the additional control options available to select from.**

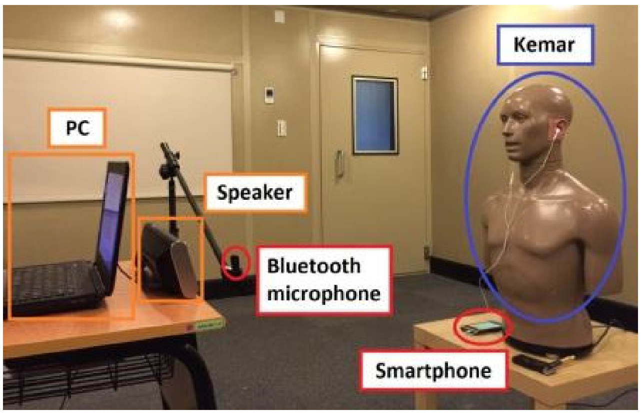


FIGURE 9. **Experimental setup for evaluation of SmartHear voice transmission-based functions and the voice-to-text function. The Kemar is not used as part of the voice-to-text evaluation.**

scenarios are compared. As depicted in Fig. 9, the "Listen" mode is adopted in the experiment to reflect typical usage. The "Record" option is not explicitly evaluated as part of the experiment because it does not trigger any extra technical audio processing other than the capability to record.

The hardware details are summarized in Table 3. The particular microphone selected in our experiment is a clip-on headset that can conveniently be equipped by a teacher in an actual classroom setting. Moreover, the Kemar's ear is set up with a G.R.A.S. RA0045 Polarized Ear Simulator installed inside its G.R.A.S. KB0060 Kemar right pinna. The ear simulator is designed such that the acoustic input impedance is highly similar to that of the human ear [36].

For the result calculation, we incorporate four typical audiograms ranging from normal hearing (NH) to mild hearing loss (HL) to represent the hearing condition of typical target users of SmartHear, as shown in Fig. 10. Audiograms 1 and 2 represent NH, while Audiograms 3 and 4 represent mild HL. The audiograms are referenced from among the most common audiometric configurations for a mix of men and women [37]. Since the configuration settings in [37]

TABLE 3. **Hardware used in the experiment.**

Hardware	Brand/Model
Bluetooth headset	Jabra Clipper
Smartphone	HTC One (M8)
Earphones	Apple EarPod
Kemar microphone	G.R.A.S. RA0045 Polarized Ear Simulator
Kemar ear	G.R.A.S. KB0060 Kemar Right Pinna

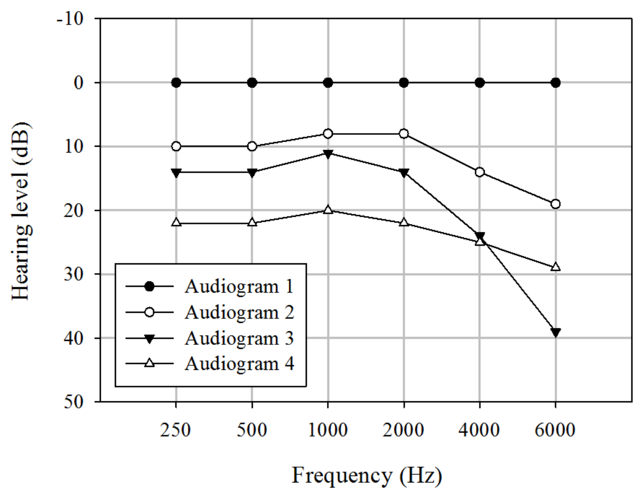


FIGURE 10. **Four typical audiograms ranging from NH to mild HL used in the evaluation of SmartHear voice transmission-based functions.**

are limited in the range of 500–8000 Hz, for the purpose of our evaluation, the same value for 500 Hz is extrapolated to 250 Hz. This yields four audiogram configurations at frequencies of 250, 500, 1000, 2000, 4000, and 6000 Hz.

The experimental procedure is summarized as follows:

1) Play and record 20 IEEE sentences inside the setup environment. These serve as the clean speech that will act as the basis of comparison for the evaluations.
2) Insert speech-shaped noise (SSN) into the clean speech to generate noisy speeches. The SNRs used

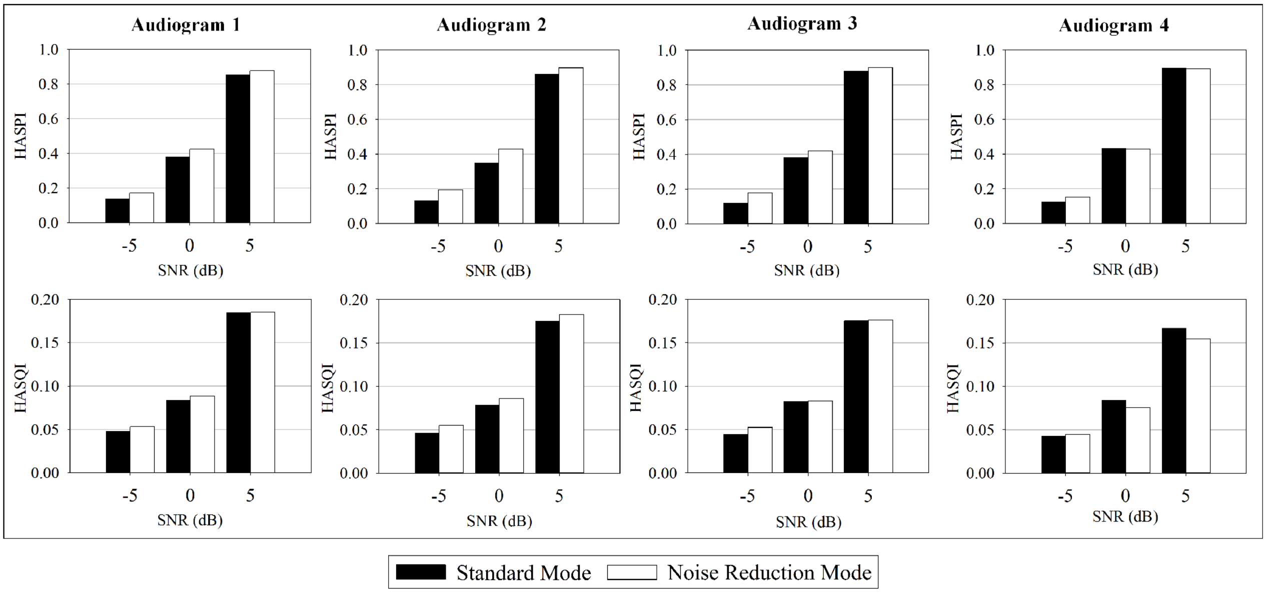


FIGURE 11. The HASPI (top four) and HASQI (bottom four) scores for three SNR conditions (−5, 0, and 5 dB) and four audiograms (Fig. 10). "Standard mode" and "noise reduction mode" refer to turning SmartHear's "Reduce Noise" option off and on, respectively. Higher HASPI and HASQI scores indicate better speech intelligibility and quality, respectively.

for this experiment include −5, 0, and 5 dB, which closely represent the typical SNRs in classroom environments [6].

3) The noisy speeches are played from the speaker, transmitted to the smartphone, and recorded by the microphone inside the Kemar's ear through earphones. This step is performed separately for the standard mode and noise reduction mode.
4) The recorded audio is fed into the HASPI/HASQI calculation program along with the corresponding original clean utterances. The output is the objective scores of all the speeches for the standard mode and noise reduction mode at the specified SNRs.

B. SmartHear's VOICE-TO-TEXT CONVERSION

SmartHear's voice-to-text (speech recognition) function is evaluated by its interpretation accuracy. The experimental setup is identical to that of the previous evaluation as shown in Fig. 9, except that the Kemar is not used here. Our experiment is performed using Arabic numerals from the TIDIGITS corpus [38]. The use of digits for speech recognition avoids domain-specific and possibly biased results due to any particular topic/context of the speech. The database contains 308 sets of speech produced by two male and two female speakers (each producing 77 sets of speech or digit sequences). All the sequences from the database are combinations of single digits (i.e., 0–9) pronounced in English. The digit 0 may legitimately be pronounced as "zero" or "oh." Each sequence contains one to seven digits. The appearances of each digit across all sequences are roughly equiprobable to avoid bias in favor of any particular digit, with the exception that the digit 0 appears twice as likely as other digits because it is pronounced in two different ways.

The recognition accuracy is calculated by dividing the number of correctly recognized sequences by the total number of sequences. A sequence is considered correctly recognized if all the digits in the sequence are correctly recognized. A digit is considered correctly recognized if the displayed text reads the same as how the digit is pronounced (e.g., for the digit 4, the displayed text of "four," "4," and "for" are all considered correct, whereas "Ford" and "fur" are considered incorrect even though they may sound similar).

This experiment compares the recognition accuracy with or without SmartHear's voice-to-text function. The latter is achieved by implementing a separate but similar Android application as SmartHear. This purpose-built Android application performs speech recognition in an identical manner as SmartHear's voice-to-text function, yet without a paired Bluetooth microphone. The smartphone is placed at distances of 5, 50, 100, 150, 200, 250, and 300 cm away from the speaker. These distances represent typical talker-listener distances in lecture or discussion scenarios. For the case with SmartHear, the Bluetooth microphone is placed at a constant 5 cm away from the speaker. 308 sequences are played one-by-one from the speaker, and their recognition results are recorded.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. PERFORMANCE OF SmartHear's VOICE TRANSMISSION

Fig. 11 shows the HASPI and HASQI scores for three SNRs (−5, 0, and 5 dB) and four audiograms (Fig. 10). As can be seen, a higher SNR consistently results in a better score throughout the four audiograms, showing the negative effects of noise on speech intelligibility and quality. A comparison between HASPI and HASQI reveals that the HASPI scores increase rapidly and towards the maximum value of 1 as the SNR increases, while the HASQI scores increase only moderately and do not exceed 0.2. The suboptimal sound quality of the audio may be attributed to the relatively low audio sampling rate of 11.025 kHz in SmartHear (thus, audio signals above half the sampling rate will be truncated), as well as the Bluetooth specifications [39] (i.e., audio signals at the transmitter should be below 4 kHz and audio signals above 4 kHz should be 20 dB below the maximum in the 0–4-kHz range at the receiver). However, as shown in this figure, sound quality does not dictate speech perceptibility/intelligibility.

A comparison between the standard mode and noise reduction mode shows that, for Audiograms 1–3, the noise reduction mode outperforms the standard mode in both HASPI and HASQI scores, with diminishing advantages as the SNR increases. This suggests that, at higher SNRs, since the original speech is quite clean already, additional signal manipulations such as noise-reduction processing may create unnatural artifacts to the speech, which outweigh its benefits. This is further validated in Audiogram 4, where the standard mode slightly outperforms the noise reduction mode at SNR = 0 dB and 5 dB, with a bigger gap in the HASQI scores because the unnatural artifacts affect especially the quality of the speech. Furthermore, the different trend for Audiogram 4 as compared to Audiograms 1–3 suggests a possible effect of the severity of hearing loss on the efficacy of the noise-reduction processing. As discussed, signal manipulations may create unnatural artifacts to the speech that affect speech intelligibility and quality, and such effects may be more pronounced for people with hearing loss.

The results carry implications for the most suitable application scenarios for the two operation modes. The standard mode is recommended for normal usage when the noise level of the surrounding environment is controlled or relatively low. This configuration exhibits the least amount of audio data manipulations and conserves the battery power. When the noise level of the surrounding environment heightens, the noise reduction mode may be used to enhance the intelligibility and quality of the speech.

B. PERFORMANCE OF SmartHear's VOICE-TO-TEXT CONVERSION

Fig. 12 compares the percent accuracy result for speech recognition with or without SmartHear. It is seen that the accuracy performance without SmartHear exhibits a steep decay with an increasing talker-listener distance, while the accuracy performance with SmartHear remains unaffected by the talker-listener distance. This shows that the talker-listener distance affects the audio reception and recognition at the listener, and SmartHear can significantly mitigate this distance effect by directly transmitting the audio to the listener. When the talker-listener distance is very small (say, less than 50 cm), SmartHear does not present an advantage. This is because the direct reception of audio signals at the smartphone microphone is sufficiently strong so that relaying the audio signals through the Bluetooth device, which might cause slight signal impairments as mentioned previously, is unnecessary and unbeneficial.

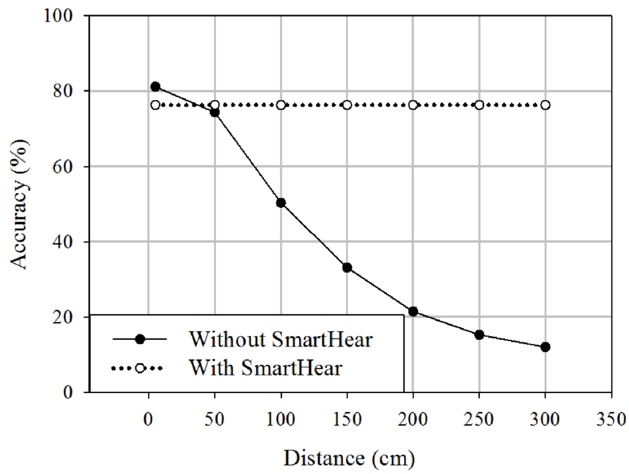


FIGURE 12. The speech recognition performance with and without SmartHear. The smartphone (audio receiver) is placed at distances of 5, 50, 100, 150, 200, 250, and 300 cm away from the speaker (audio source).

Speech converted to text, or captioning, has been proven to be an effective accommodation for students with hearing impairment to enhance their access to verbal communication in typical classrooms. Interact-AS is one of the exemplary products used in the United States, while the voice-to-text function implemented in SmartHear is unique in its application of a Bluetooth microphone placed near the talker to eliminate the negative effects of reverberation and distance on the received audio information and to enhance the performance of ASR.

C. FUTURE APPLICATIONS OF SmartHear

The proposed SmartHear system, as described earlier, finds many immediate benefits in classroom learning scenarios through its key functions such as configurable transmitter/receiver assignment, advanced noise reduction, audio recording, and voice-to-text conversion. These functions collectively facilitate personalized learning for general students and students with special needs alike, by offering personalized benefits such as improved SNR of the received speech signal at the listener, easy recording of the audio lectures for future reference or playback, and reduced listening effort with visual/language aids. SmartHear distinguishes itself from most existing classroom learning enhancement approaches in that SmartHear targets specifically at students with special needs or disabilities to facilitate their learning [9], [40], [41], with its broad applicability extensible to general students as well.

The proposed SmartHear system (and its future extension) could also find immense potential in the online learning paradigm. SmartHear is built upon a smartphone platform and therefore could be a handy online learning facilitator. For instance, the recorded lectures and the corresponding transcripts using SmartHear can be made available online for future reference. SmartHear could also provide a platform for forming social learning networks in the online learning community. In a possible future extension of SmartHear system featuring embedded sensors, location-, situation-, and time-dependent information could be provided for monitoring a student's engagement, behavior, and performance in online learning. As smartphone technologies are constantly developing, the possibilities are unlimited. In summary, prospects for greater applicabilities of SmartHear, with its core framework established on smartphones and its core functions targeting at students with special needs and beyond, remain bright in the future human learning paradigm.

VI. CONCLUSION

We have developed a smartphone-based multi-functional hearing assistive system (SmartHear) to facilitate classroom learning. The proposed system is a highly affordable and accessible option for a variety of individuals who could benefit from a better listening experience in a learning environment, including: students with hearing loss, students with ADHD, students in the foreign language classroom, etc. SmartHear incorporates many important functions to enhance listening experiences in the classroom: 1) configurable transmitter/receiver assignment, to allow flexible designation of transmitter/receiver roles; 2) advanced noise-reduction techniques, to mitigate the background noise effect on the voice signals; 3) audio recording, to enable saving an audio copy on the smartphone for future playback or reference; and 4) voice-to-text conversion, to give students visual text aid which is particularly useful in small group discussions. Experimental results based on standard speech perception (HASPI) and quality (HASQI) testing suggested using the standard mode (i.e., turning off the ''Reduce Noise'' option) in higher-SNR environments to avoid unnecessary signal manipulations and conserve power, and using the noise reduction mode in lower-SNR environments to reduce the background noise at the cost of higher battery power consumption. The voice-to-text evaluation experiment showed SmartHear's ability in maintaining voice-to-text conversion accuracy regardless of the distance between the speaker and listener. Future work includes enhancing the current voice-to-text function to achieve continuous transcription and/or text dissemination among multiple smartphones via WiFi, adding new application features to further improve the classroom learning experience for users (e.g., concurrent audio and visual aid), and conducting experiments with actual students inside a classroom to subjectively evaluate the SmartHear functions.

ACKNOWLEDGMENT

The authors gratefully acknowledge Kylon Chiang and Preston So for their respective contributions in designing the graphical user interface and the multimedia video for SmartHear.

REFERENCES

[1] R. Emanuel *et al.*, ''How college students spend their time communicating,'' *Int. J. Listening*, vol. 22, no. 1, pp. 13–28, 2008.

[2] F. H. Bess and B. W. Hornsby, ''Commentary: Listening can be exhausting–fatigue in children and adults with hearing loss,'' *Ear Hearing*, vol. 35, no. 6, pp. 592–599, Nov./Dec. 2014.

[3] P. A. Gosselin and J. P. Gagne, ''Older adults expend more listening effort than young adults recognizing speech in noise,'' *J. Speech, Lang., Hearing Res.*, vol. 54, no. 3, pp. 944–958, Nov. 2011.

[4] J. L. Desjardins and K. A. Doherty, ''Age-related changes in listening effort for various types of masker noises,'' *Ear Hearing*, vol. 34, no. 3, pp. 261–272, May/Jun. 2013.

[5] C. B. Hicks and A. M. Tharpe, ''Listening effort and fatigue in school-age children with and without hearing loss,'' *J. Speech, Lang. Hearing Res.*, vol. 45, no. 3, pp. 573–584, Jun. 2002.

[6] C. S. Howard, K. J. Munro, and C. J. Plack, ''Listening effort at signal-to-noise ratios that are typical of the school classroom,'' *Int. J. Audiol.*, vol. 49, no. 12, pp. 928–932, Dec. 2010.

[7] B. W. Hornsby, ''The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands,'' *Ear Hearing*, vol. 34, no. 5, pp. 523–534, Sep. 2013.

[8] S. Fraser, J.-P. Gagné, M. Alepins, and P. Dubois, ''Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues,'' *J. Speech, Lang., Hearing Res.*, vol. 53, no. 1, pp. 18–33, Feb. 2010.

[9] M. S. Stinson, L. B. Elliot, R. R. Kelly, and Y. Liu, ''Deaf and hard-of-hearing students' memory of lectures with speech-to-text and interpreting/note taking services,'' *J. Special Edu.*, vol. 43, no. 1, pp. 52–64, May 2009.

[10] M. A. Gernsbacher, ''Video captions benefit everyone,'' *Policy Insights Behavioral Brain Sci.*, vol. 2, no. 1, pp. 195–202, Oct. 2015.

[11] M. Stinson, ''Current and future technologies in the education of deaf students,'' in *The Oxford Handbook of Deaf Studies, Language, and Education*, vol. 2. New York, NY, USA: Oxford Univ. Press, 2010, pp. 93–100.

[12] *Acoustical Performance Criteria, Design Requirements and Guidelines for Schools*, Amer. Nat. Standards Inst., Washington, DC, USA, document 12.60-2002, 2002.

[13] *American Speech-Language-Hearing Association (ASHA) Ad Hoc Committee on FM Systems, Guidelines for Fitting and Monitoring FM Systems*, accessed on May 2, 2017. [Online]. Available: http://www.asha.org/policy/GL2002-00010

[14] E. C. Schafer and M. P. Kleineck, ''Improvements in speech recognition using cochlear implants and three types of FM systems: A meta-analytic approach,'' *J. Educ. Audiol.*, vol. 15, pp. 4–14, 2009.

[15] L. Thibodeau, ''Benefits of adaptive FM systems on speech recognition in noise for listeners who use hearing aids,'' *Amer. J. Audiol.*, vol. 19, no. 1, pp. 36–45, 2010.

[16] J. Wolfe, E. C. Schafer, B. Heldner, H. Mülder, E. Ward, and B. Vincent, ''Evaluation of speech recognition in noise with cochlear implants and dynamic FM,'' *J. Amer. Acad. Audiol.*, vol. 20, no. 7, pp. 409–421, 2009.

[17] K. L. Anderson, H. Goldstein, L. Colodzin, and F. Iglehart, ''Benefit of S/N enhancing devices to speech perception of children listening in a typical classroom with hearing aids or a cochlear implant,'' *J. Educ. Audiol.*, vol. 12, pp. 16–30, 2005.

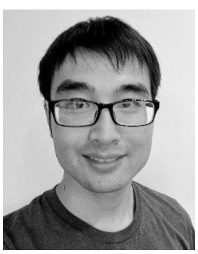
[18] O. A. Lopez, E. A. amd Costa, and D. V. Ferrari, ''Development and technical validation of the mobile based assistive listening system: A smartphone-based remote microphone,'' *Amer. J. Audiol.*, vol. 25, pp. 288–294, Oct. 2016.

[19] *Jacoti Lola Classroom*, accessed on Jun. 2017. [Online]. Available: https://www.jacoti.com/lola

[20] J. Hornickel, S. G. Zecker, A. R. Bradlow, and N. Kraus, ''Assistive listening devices drive neuroplasticity in children with dyslexia,'' *Proc. Nat. Acad. Sci.*, vol. 109, no. 41, pp. 16731–16736, 2012.

[21] N. Koohi, D. Vickers, J. Warren, D. Werring, and D.-E. Bamiou, ''Long-term use benefits of personal frequency-modulated systems for speech in noise perception in patients with stroke with auditory processing deficits: A non-randomised controlled trial study,'' *BMJ Open*, vol. 7, no. 4, p. e013003, 2017.

[22] E. C. Schafer *et al.*, "Personal FM systems for children with autism spectrum disorders (ASD) and/or attention-deficit hyperactivity disorder (ADHD): An initial investigation," *J. Commun. Disorders*, vol. 46, no. 1, pp. 30–52, 2013.
[23] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Heidelberg, Germany: Springer, 2005.
[24] P. C. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL, USA: CRC Press, 2013.
[25] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 1, pp. 7–19, Jan. 2015.
[26] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Proc. Interspeech*, Aug. 2013, pp. 436–440.
[27] A. Narayanan and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 7092–7096.
[28] Y. Tsao and Y.-H. Lai, "Generalized maximum a posteriori spectral amplitude estimation for speech enhancement," *Speech Commun.*, vol. 76, pp. 112–126, Feb. 2016.
[29] *Multimedia Programming Interface and Data Specifications 1.0*, IBM Corp. Microsoft Corp., Aug. 1991.
[30] Android Developers. *Speech Recognizer*, accessed on Jun. 2017. [Online]. Available: http://developer.android.com/reference/android/speech/SpeechRecognizer.html
[31] J. M. Kates and K. H. Arehart, "The hearing-aid speech perception index (HASPI)," *Speech Commun.*, vol. 65, pp. 75–93, Nov./Dec. 2014.
[32] J. M. Kates and K. H. Arehart, "The hearing-aid speech quality index (HASQI)," *J. Audio Eng. Soc.*, vol. 62, no. 3, pp. 99–117, Mar. 2010.
[33] *Jabra Clipper Bluetooth*, accessed on Jun. 6, 2017. [Online]. Available: http://www.jabra.com/supportpages/jabra-clipper#/#100-96800000-02
[34] *HTC One M8*, accessed on Jun. 6, 2017. [Online]. Available: http://www.htc.com/us/smartphones/htc-one-m8/
[35] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *J. Acoust. Soc. Amer.*, vol. 97, no. 6, pp. 3907–3908, Jun. 1995.
[36] *G.R.A.S. Sound & Vibration, G.R.A.S. RA0045 Externally Polarized Ear Simulator According to IEC 60318-4 (60711)*, accessed on Jun. 2017. [Online]. Available: http://www.gras.dk/ra0045.html
[37] L. Ciletti and G. A. Flamme, "Prevalence of hearing impairment by gender and audiometric configuration: Results from the National Health and Nutrition Examination Survey (1999-2004) and the Keokuk County Rural Health Study (1994-1998)," *J. Amer. Acad. Audiol.*, vol. 19, no. 9, pp. 672–685, Oct. 2008.
[38] R. G. Leonard, "A database for speaker-independent digit recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 1984, pp. 328–331.
[39] Bluetooth Special Interest Group. (2001). *Specification of the Bluetooth System, Version 1.1*. [Online]. Available: http://www.bluetooth.com
[40] T. S. Hasselbring and C. H. W. Glaser, "Use of computer technology to help students with special needs," *Future Children*, vol. 10, no. 2, pp. 102–122, Oct. 2000.
[41] A. Magnan and J. Ecalle, "Audio-visual training in children with reading disabilities," *Comput. Edu.*, vol. 46, no. 4, pp. 407–425, 2006.



ALAN CHERN received the B.S. degree in chemical and biomolecular engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011. He was a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan, in 2016. He is currently pursuing the M.S. degree with the School of Computer Science, Georgia Institute of Technology. His main research interests include software engineering and mobile application development.

YING-HUI LAI (M'16) received the B.S. degree in electrical engineering from National Taiwan Normal University, Taipei, Taiwan, in 2005, and the Ph.D. degree in biomedical engineering from National Yang-Ming University, Taipei, in 2013. From 2010 to 2012, he was involved in the development of hearing aid with Aescu Technologies. From 2013 to 2016, he was a Post-Doctoral Research Fellow with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. Since 2016, he has been with the Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan, where he is currently an Assistant Professor. His research interests include hearing aids, cochlear implants, noise reduction, pattern recognition, and source separation.

YI-PING CHANG received the B.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2000, the M.S. degree in electrical and control engineering from National Chiao Tung University, Hsinchu, in 2002, and the Ph.D. degree in biomedical engineering from the University of Southern California, Los Angeles, CA, USA, in 2009. She also received the Graduate Certificate in educational studies with a specialization in listening and spoken language from the University of Newcastle, Australia, in 2014. From 2009 to 2010, she was a Post-Doctoral Researcher with the House Ear Institute, Los Angeles. Since 2011, she has been with the Children's Hearing Foundation (CHF), Taipei, Taiwan, where she is currently the Director of CHF's Speech and Hearing Science Research Institute. She has been an Adjunct Assistant Professor with the Department of Audiology and Speech-Language Pathology, Mackay Medical College, New Taipei, since 2015. Her research interests include speech perception in cochlear implants, bimodal hearing, and assessment of listening and spoken language development of children with hearing loss.

YU TSAO (M'09) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology in 2008. From 2009 to 2011, He was a Researcher with the National Institute of Information and Communications Technology, Japan, where he was involved in research and product development in automatic speech recognition for multilingual speech-to-speech translation. He is currently an Associate Research Fellow with the Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan. His research interests include speech and speaker recognition, acoustic and language modeling, audio-coding, and bio-signal processing. He received the Academia Sinica Career Development Award in 2017.

RONALD Y. CHANG (M'12) received the B.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2000, the M.S. degree in electronics engineering from National Chiao Tung University, Hsinchu, in 2002, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2008. From 2002 to 2003, he was with the Industrial Technology Research Institute, Hsinchu. In 2008, he was a Research Intern with the Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. In 2009, he was involved with the NASA Small Business Innovation Research projects. Since 2010, he has been with the Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan, where he is currently an Assistant Research Fellow. His research interests include MIMO communications and interference networks, energy-aware communications, Internet of Things, and e-health/m-health. He was a recipient of the Best Paper Award from the IEEE Wireless Communications and Networking Conference in 2012, and the Outstanding Young Scholar Award from the Ministry of Science and Technology, Taiwan, in 2015. He was an Exemplary Reviewer for the IEEE Communications Letters in 2012 and the IEEE Transactions on Communications in 2015.

HSIU-WEN CHANG received the first master's degree in linguistics from National Chung Cheng University, Chiayi, Taiwan, in 1997, the second master's degree in audiology from Washington University in St. Louis, St. Louis, MO, USA, in 2001, and the Ph.D. degree in biomedical engineering from National Yang-Ming University, Taipei, Taiwan in 2012. She is a certified audiologist and has extensive clinical experience in adult and pediatric hearing services. She is currently an Assistant Professor with the Department of Audiology and Speech-Language Pathology, Mackay Medical College, New Taipei, Taiwan. She performs research into many aspects of audiology. Her research interests include the mobile app development for hearing aids, the electrophysiological evaluation of hearing aid effectiveness in infants and young children, and the development of outcome assessment tools in mandarin chinese. She also serves as a Board Member of the International Association of Logopedics and Phoniatrics.