

Article

Not peer-reviewed version

Using Voice-to-Text Transcription to Examine Outcomes of AirPods Pro Receivers When Used as Part of a Remote Micro-Phone System

[Shuang Qi](#) * and [Linda Thibodeau](#)

Posted Date: 19 March 2025

doi: 10.20944/preprints202503.1408.v1

Keywords: remote microphone system; voice-to-text transcription; AirPods Pro; typical hearing



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Using Voice-to-Text Transcription to Examine Outcomes of AirPods Pro Receivers When Used as Part of a Remote Micro-Phone System

Shuang Qi ^{1,2,*} and Linda Thibodeau ^{1,2}

¹ Callier Center for Communication Disorders; e-mail@e-mail.com

² Department of Speech, Language and Hearing Sciences, The University of Texas at Dallas; e-mail@e-mail.com

* Correspondence: shuang.qi@utdallas.edu

Featured Application: Smartphone-based remote microphone systems may provide speech recognition benefits in noisy environments.

Abstract: Hearing difficulty in noise can occur to 10-15% of listeners with typical hearing in the general population of the United States. Using one's smartphone as a remote microphone (RM) system with AirPods Pro (AP) may be considered as an assistive device given its wide availability and possible lower price. To evaluate this possibility, the accuracy of voice-to-text transcription for sentences presented in noise was compared when KEMAR wore an AP receiver connected to an iPhone set to function as an RM system to the accuracy obtained when he wore a sophisticated Phonak Roger RM system. A ten-sentence list was presented for six technology arrangements at three signal-to-noise ratios (SNR; +5, 0, and -5 dB) in two types of noise (speech-shaped and babble noise). Each sentence was transcribed by Otter AI to obtain an overall percent accuracy for each condition. At the most challenging SNR (-5 dB SNR) across both noise types, the Roger system and smartphone/AP set to noise cancellation mode showed significantly higher accuracy relative to the condition when smartphone/AP was in transparency mode. However, the major limitation of Bluetooth signal delay when using the AP/smartphone system would require further investigation in real-world settings with human users.

Keywords: remote microphone system; voice-to-text transcription; AirPods Pro; typical hearing

1. Introduction

Noise is common in everyday life, affecting communication not only for individuals with hearing impairments but also for those with typical hearing (TH), i.e. bilateral pure-tone thresholds ≤ 20 dB HL for 0.5-8 kHz in octave bands. Although pure-tone audiometry is widely recognized as the "gold standard" for assessing hearing function, a normal audiogram does not necessarily guarantee satisfactory communication experiences in all settings. Over 10% of adults may experience hearing difficulty in noisy environments despite having normal audiograms. Tremblay et al. [1] reported that 12% of 682 individuals aged 21-67 years with TH reported hearing difficulties. Similarly, 15% of 2015 subjects aged 20-69 years with bilateral pure-tone averages of ≤ 25 dB HL for 0.5, 1, 2, and 4 kHz experienced similar issues [2]. However, standard clinical guidelines from the American Academy of Audiology [3] and the American Speech-Language-Hearing Association [4] provide limited guidance for addressing these communication challenges for listeners with TH, aside from recommendations to use hearing protectors to prevent occupational noise-induced hearing loss [5]. Mealings et al. [6] found that many clients express frustration when told they have "normal hearing" while still experiencing hearing difficulties. Additionally, clinicians expressed frustration with lacking adequate training or scientific resources to assist these clients beyond suggesting the use of communication strategies [6].

One potential technological solution is to use remote microphone (RM) systems, designed to enhance the signal-to-noise ratio (SNR). These systems consist of a transmitter/microphone worn by the speaker and a receiver worn by the listener [7]. Research has shown that RM systems significantly improve speech recognition in noisy environments for individuals with hearing impairments and typical hearing. Thibodeau [8] reported that the Phonak Roger RM system enhanced speech recognition performance in persons with moderate-to-severe hearing loss by 45% to 61% compared to using hearing aids (HAs) or cochlear implants alone. In 2024, Thibodeau et al. [9] reported that RM systems benefited adults with TH. The observed improvements were 21.37% and 9.87% when using the Phonak Roger Select and Roger Pen, respectively, compared to no RM system in 10 participants with TH, aged 20 to 63 years [9]. Shiels and colleagues [10] found RM benefits for children, ages 6 to 12 years. They reported improvements in the sentence recognition scores by an average of 3.03, 21.40, 21.61, and 49.95% when tested in SNRs of +12, +5, 0, and -10 dB, respectively. They also reported improvements in visual and auditory attention with RM use.

Although listeners with TH can potentially benefit from using RM systems designed for persons with hearing loss [9] the adoption of such devices is limited due to their high cost. For example, Mealings et al. [11] reported none of the participants ($n = 27$) with TH expressed a preference for purchasing a pair of mild-gain HAs priced at approximately \$3250, despite the potential benefits for their hearing difficulties. Additionally, the higher expense associated with RM systems and the inconvenience of carrying multiple devices may deter individuals from using RM systems in daily communication. Furthermore, the social stigma often associated with using hearing devices can be a concern for some people. da Silva et al [12] reported that persons with hearing impairment might be perceived as “less intelligent” or labeled as “deaf and dumb,” causing feelings of shame and embarrassment. These negative emotional feelings of self-imposed stigma may lead to social isolation and affect decision-making regarding the initial acceptance and pursuit of potential treatment [12]. Therefore, it is essential to find more affordable and approachable alternatives to increase the acceptance and use of hearing devices for addressing hearing difficulties.

As an alternative, smartphones, tablets, and laptop computers equipped with built-in microphones and Bluetooth capabilities (Bluetooth Special Interest Group, Kirkland, WA) have been suggested as components of RM systems for use in small-group conversations for adults. Although not designed to be worn at an optimal distance from the talker’s mouth, these devices do not require external components or additional hardware [13]. Furthermore, smartphones are readily used, with nearly everyone carrying one throughout the day. It has been reported that 98% of Americans own a cellphone or smartphone [14]. Moreover, Bluetooth-compatible headphones for iOS (Apple Inc., Cupertino, CA) and Android (Open Handset Alliance, Mountain View, CA) operating systems have been introduced, making it easier to integrate RM system functionality directly through smartphones [15]. However, there are limitations to using Bluetooth receivers as part of an RM system considering their potential 30 to 100 milliseconds of audio transmission delay [7].

One smartphone-based RM system involves the iPhone and AirPods Pro (AP) earphones (Apple Inc., Cupertino, CA). It is noteworthy that the software available on compatible versions of AP was recently approved by the U.S. Food and Drug Administration in 2024 as an over-the-counter HA for adults with mild-to-moderate hearing loss [16]. The AP can connect via Bluetooth with iPhones and Android phones and serve as receivers worn by the listener, with the smartphone functioning as a transmitter/microphone when placed close to the talker. In Android phones (i.e., Samsung, Google Pixel), the RM features that can be used for enhancing hearing perception in noise include ‘Hearing Enhancement’ in Samsung series, and “Sound Amplifier” in Google Pixel. In iPhone devices, Live Listen (LL) is an additional native feature, available since 2014, that enables microphone transmission to compatible hearing devices such as HA or AP headphones with Bluetooth Low Energy protocols. When activated, LL uses the iPhone’s microphone to capture sound and transmit it wirelessly from the iPhone to the receiver (i.e., AP), amplifying and clarifying the sounds in their immediate environment for users [17]. AP can be set at three modes: transparency (TP; lets outside sound in), noise cancellation (NC; canceling the external sounds), and off. This could potentially be a

lower-cost alternative to a dedicated RM system designed for persons with TH. One such RM system, a Phonak Roger Touchscreen transmitter used with a Roger Focus II receiver, was shown to enhance speech intelligibility by an average of 53% in 16 children (ages 8-16 years) with unilateral hearing loss, compared to their peers with TH [18].

Several studies have highlighted the potential of using AP receivers for adults with or without hearing impairment [19–24], as summarized in chronological order in Table 1. For instance, Lin et al. [19] verified that AP receivers have comparable electroacoustic results to HAs, providing adequate amplification for individuals with mild-to-moderate hearing loss. Hammond and Diedesch [20] found that listeners appreciated the custom audiogram-driven features in AP, finding them easy to use and beneficial regardless of hearing status. Valderrama et al. [22] reported that using AP and an iPhone mitigated hearing challenges for individuals with TH, resulting in a significant 11.8% increase in speech intelligibility and a +5.5 dB SNR advantage compared to baseline conditions without AP (unaided). Only one master's thesis by Foroogozar [24] compared the use of LL versus non-use, showing significant improvements in memory retention from 43.8% (no LL) to 59.4% (with LL) and mean sentence recognition scores from 81.8% (no LL) to 94.4% (with LL) in 23 adults aged 60 and above with normal to mild/moderate hearing loss.

Considering the focus of this study was to evaluate the AP when used with a smartphone set to LL compared to a dedicated RM system, it is essential to provide objective verification and limit potential confounding human factors (e.g., age, gender, education, personality, attention, linguistic familiarity, cognition). An objective, non-biased speech recognition method was developed following the COVID-19 pandemic which disrupted traditional methods of collecting experimental data involving human subjects. This new approach was developed utilizing voice-to-text transcription (VTT), the Knowles Electronics Manikin for Acoustic Research (KEMAR) with a standardized artificial ear (Zwislocki coupler), to replace human responses. Advancements in artificial intelligence have led to the creation of various transcription tools designed to convert spoken words into text with great accuracy as a way to facilitate closed captioning [25]. Such VTT applications include Otter.ai (<https://otter.ai/>), Sonix (<https://sonix.ai>), Trint (<https://trint.com/>), and Google Cloud Speech-to-Text (<https://cloud.google.com/speech-to-text/>). In a comparative study assessing various VTT tools, Otter.ai was reported to provide the most accurate transcripts with an accuracy rate of 99.7% [26]. Given this high accuracy rate, Otter.ai was utilized for VTT in this study.

The aim of this study was to compare the VTT accuracy of two RM systems (the iPhone set to LL and the Roger RM), specifically designed for users with TH. The investigation involved three SNRs, one HINT list, and two types of noise. The research question was: How does the transcription accuracy for AP in TP and NC modes when used with an iPhone set to LL compared to the Roger RM system across two noise types and three SNR levels? It was hypothesized that transcription accuracy would be superior in speech noise compared to babble noise, for the higher SNR conditions, and with the Roger RM system relative to the smartphone-based RM system.

Table 1. A summary of studies using AirPods (Pro) as a hearing assistive device in chronological order.

Study	Participants N and Age	Hearing	Methods Devices	Tests	Findings
Lin et al. (2022) [19]	<ul style="list-style-type: none"> N= 21 Mean = 42.9 years 	Mild-to-moderate hearing loss	<ol style="list-style-type: none"> 1. iPhone XS 2. AirPods 2 3. AP 4. A premium HA (OTICON Opn 1) 5. A basic HA (Bernafon MD 1) 	<ul style="list-style-type: none"> • Electroacoustic analysis • SRT and MHINT in quiet and in noise, aided versus unaided 	<ul style="list-style-type: none"> • AP has comparable electroacoustic results to HAs and thus can provide proper amplification for listeners with mild-to-moderate hearing loss, while AirPods 2 does not. • SRT and MHINT scores varied with different directions of the noise source. AP worked similarly to HAs when tested in easier discriminating conditions (noise from another direction) while significantly poorer when tested in harder conditions (speech and noise came in the same direction)
Hammond & Diedesch (2023) [20]	<ul style="list-style-type: none"> N = 25 24 in 18-25 years. one in 46-60 years 	7 with self-reported hearing loss and 18 with self-reported TH	<ol style="list-style-type: none"> 1. Participants' personal iPhone 2. AP 	<ul style="list-style-type: none"> • Mimi Hearing Test • AzBio sentences in quiet and noise (+5 dB SNR) • Sound quality ratings • A ten-item questionnaire on usability, benefit, and satisfaction. 	<ul style="list-style-type: none"> • Usability: Participants found the custom features easy to use. All participants felt confident in using and teaching others to use these features regardless of their hearing status. • Listening Experience: There was variability in satisfaction with the custom features. Hearing-impaired participants were more likely to recommend and continue using the features than TH participants. Hearing-impaired participants rated the features in noisy environments more positively than TH participants. • Perceived Benefit: Overall, participants found the custom features more beneficial in quiet than in noise. Hearing-impaired participants reported greater benefit from the features in both quiet and noisy conditions compared to TH participants.
Martinez (2023) [21]	None		<ol style="list-style-type: none"> 1. Head simulator 2. AP 3. AP 2 4. iPhone 	Noise attenuation of AP and AP 2 in different modes (NC, TP, and Off) under various noise conditions at different levels.	<ul style="list-style-type: none"> • NC mode: AP 2 reduced ambient noise by an average of 27 dB across all frequencies, showing a 12 dB improvement in low-frequency noise reduction compared to the AP. • TP Mode: The new adaptive TP feature in AP 2 lowered loud sounds over 85 dBA, ensuring safe listening while maintaining awareness of the environment.
Valderrama et al. (2024) [22]	<ul style="list-style-type: none"> N = 17 Ages 21-59 	TH but self-reported hearing difficulties in noise	<ol style="list-style-type: none"> 1. iPhone 2. AP 	<ul style="list-style-type: none"> • AP set with custom headphone accommodations, including TP, Conversation Boost and Ambient Noise Reduction. • Real-ear measures. • Speech-in-noise intelligibility tests using the BKB sentences • Subjective listening effort evaluation 	<ul style="list-style-type: none"> • CB and the noise cancellation feature provided +5.4 dB SNR advantage, which is comparable to about 3-6 dB SNR benefit provided by directional HA microphones. • AP significantly improved speech intelligibility by 11.8%, from 54.6% unaided to 66.4%. Subjective listening effort evaluation suggested AP significantly reduced the mental demand by 8% and effort by 8%. • Main barriers to using AP as an assistive listening device: <ul style="list-style-type: none"> ○ Limited Hearing Benefit: 31.6% reported issues with unnatural high-frequency sound amplification, self-hearing of noises through bone conduction, and insufficient background noise attenuation. ○ Discomfort: 18.4% mentioned the AP was uncomfortable or not fitting properly. Comfort varied significantly among users.

					<ul style="list-style-type: none">• Real-world hearing experience survey	<ul style="list-style-type: none">○ Stigma and Embarrassment: 15.5% reported they felt uncomfortable wearing AP in social settings, often needing to explain their use.• Despite the barriers, around 30% indicated they would continue using AP for better hearing in noise, suggesting its potential for further enhancements.
Kim et al. (2024) [23]	<ul style="list-style-type: none">• N = 35• Median = 63 years	Mild to moderate hearing loss	<ol style="list-style-type: none">1.2.3.	iPhone AP The Bean (a PSAP)	<ul style="list-style-type: none">• Functional gain• Monosyllabic word in quiet• Sentence in noise at 0 dB SNR.• Real-ear: aided accuracy relative to NAL-NL2 targets.	<ul style="list-style-type: none">• Functional gain: No significant difference between the Bean and AP, except at 8 kHz where the Bean performed better.• Word and sentence recognition: Both Bean (median = 84% and 95%) and AP (median = 84% and 93%) significantly improved scores compared to unaided (median = 60% and 65%) conditions, with no significant difference between the Bean and AP.• Real-ear: Both devices amplified appropriately relative to the NAL-NL2 fitting formula (< 10 dB SPL between target and response frequencies across all ranges), though AP showed a deviation at 6 kHz in the left ear (12.25 dB SPL).
Foroogozar (2024) [24]	<ul style="list-style-type: none">• N = 23• Age ≥ 60	Normal to mild/moderate hearing loss	<ol style="list-style-type: none">1.2.	iPhone with LL AP2	<ul style="list-style-type: none">• A 9-digit span test to assess working memory.• QuickSIN to assess participant's SNR loss	<p>The improvement in memory retention and recognition accuracy are both significant.</p> <ul style="list-style-type: none">• Mean memory retention: 43.8% without LL and 59.4% with LL• Mean Recognition score: 81.8% without LL and 94.4% with LL <p>A positive correlation between SNR loss and recognition improvement (slope = 1.095, $R^2 = 0.284$) showed the potential of LL feature to benefit those with higher SNR loss.</p>

NOTE: AP = AirPods Pro, AP 2 = AirPods Pro second generation, SRT = speech reception threshold, MHINT = Mandarin hearing in noise test, HA = hearing aid, PSAP = personal sound amplification products, NC = Noise Cancellation, TP = Transparency, SNR = signal-to-noise ratio, BKB = Bamford-Kowal-Bench, NAL-NL2 = National Acoustics Laboratories-Non-Linear prescription, 2nd generation, LL = Live Listen, CB = Conversation Boost.

2. Materials and Methods

2.1. Equipment

To compare the accuracy of speech recognition in noise of a smartphone-based RM system (iPhone set to LL) with a dedicated RM system (Phonak Roger On with Roger Focus II), both designed for TH users, the equipment is described in Table 2 and illustrated in Figure 1.

Table 2. Devices used in the study.

Function	Name	Device
Sentence and noise stimuli input Otter VTT	Laptops	ThinkPad P70 (Lenovo)
		Latitude 7480 (Dell)
Stimuli transmission	Audiometer	Grason Stadler GSI 61
	Speakers	Two free-field corner speakers (Grason Stadler GSI)
	KEMAR	KEMAR Dummy-Head
Transmitter	Smartphone-based RM system	iPhone 12 (IOS 15.0, Apple)
Receiver		Right ear AP (1st Gen., Apple)
Transmitter	Phonak Roger RM system	Roger On
Receiver		Roger Focus II

NOTE: VTT = voice-to-text transcription, KEMAR = Knowles Electronics Manikin for Acoustic Research, RM = remote microphone, AP = AirPods Pro.

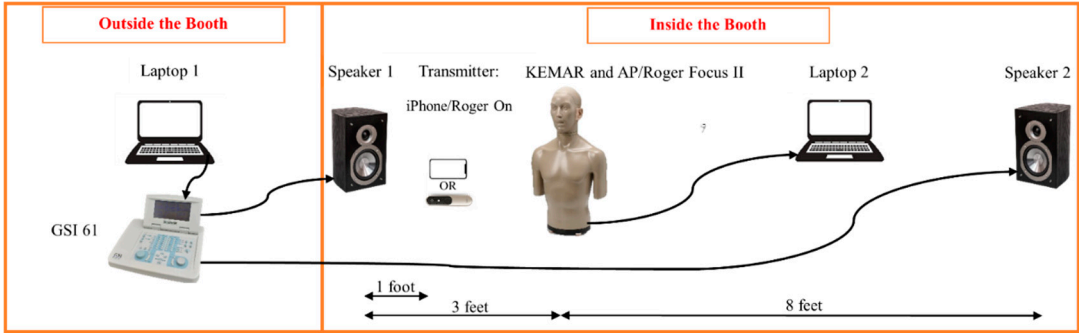


Figure 1. Equipment setup. AP = AirPods Pro, KEMAR = Knowles Electronics Manikin for Acoustic Research.

2.2. Stimuli

2.2.1. Sentences

One Hearing in Noise Test (HINT) [27] list (i.e., 20) was selected as the input, containing 10 sentences, as shown in Table 3. The 10 sentences consisted of a total of 54 words. All sentences were presented at 65 dB SPL through a speaker positioned at 0° azimuth from the KEMAR dummy head at three feet. The same list was played continuously four times for each test condition with the last three used for scoring. The sentences were generated on a Lenovo ThinkPad P70 laptop and then transmitted to channel A of the GSI-61 audiometer.

Table 3. Hearing in Noise Test list 20.

Sentence played	Scoring (points)
A/The clown has/had a/the funny fac	6
The bath water is/was warm	5
She injured four of her fingers	6
He paid his bill in full	6
They stared at a/the picture	5

A/The driver started a/the car	5
A/The truck carries fresh fruit	5
A/The bottle is/was on a/the shelf	6
The small tomatoes are/were green	5
A/The dinner plate is/was hot	5
Total	54

NOTE: The symbol "/" is used to indicate that either description is considered a correct recognition and is counted as one point towards the overall score.

2.2.2. Noise

To assess the efficacy of VTT accuracy, three SNRs were tested: +5, 0, and -5 dB, which commonly represent everyday listening scenarios. For instance, in acoustically untreated classrooms, SNRs can range from -7 to +6 dB [28]. Two noise types were employed: speech and babble noise [29], representing the most common energetic and informational masking in daily life. The noise was continuously presented at 60, 65, and 70 dB SPL (SNRs of +5, 0, and -5 dB, respectively) through a speaker located at 180° azimuth from the KEMAR dummy head at eight feet. The babble noise was generated on the Lenovo ThinkPad P70 laptop and transmitted to channel B of the GSI-61 audiometer, while the speech noise was directly produced by the GSI-61 audiometer. The order of noise was fixed for the SNR conditions but counterbalanced across the technology conditions.

2.2.3. Scoring

To transcribe the sentences played in each noise type at different SNRs, the online program Otter.ai (version 2.3.116) was used on a Dell Latitude 7480 laptop connected to the KEMAR dummy head. Each accurately transcribed word was scored one point. The total percentage of correctly transcribed words was calculated for each list, with a denominator of 54. To account for potential delays in Otter’s transcription during the initial trial, only the last three of the four presentations of the list of sentences were scored.

2.3. Procedures

2.3.1. Setup

The setup of the testing procedure is shown in Figure 1. Stimuli were presented to speakers 1 and 2 inside the sound booth, transmitted by laptop 1 (ThinkPad P70; Lenovo) and the audiometer (Grason Stadler GSI 61). Within the sound booth, speaker 1 continuously played sentence signals (HINT list 20) at 65 dB SPL, while speaker 2 was continuously playing noise signals at 60, 65, and 70 dB SPL. These signals were received by either an iPhone or Roger On through air conduction, then transmitted to the AP/Roger Focus II worn by KEMAR on the right ear. The power module integrated into KEMAR received the transmitted signal and subsequently transmitted it to laptop 2 (Latitude 7480; Dell). Finally, the sentences were transcribed using the Otter VTT program in laptop 2.

2.3.2. Test conditions

Six test conditions were categorized into three baseline settings and three Roger RM settings. The three baseline settings included no use of a smartphone: 1) AP NC: AP in Noise Cancellation Mode (NC); 2) AP TP: AP in Transparency Mode (TP), and 3) KEMAR: without any additional technological devices (i.e., no iPhone, AP, or Roger system). The three RM settings included: 1) AP NC + LL: AP set to NC while wirelessly connected to the iPhone using LL, 2) AP TP + LL: AP set to TP while wirelessly connected to the iPhone using LL, and 3) Roger: Roger On transmitter with a Roger Focus II receiver.

3. Results

The mean transcription accuracy scores for No RM conditions (Baseline) and RM conditions are presented in Tables 4 and 5 and Figure 2. There were three factors of interest including technology condition, noise type, and SNR. In general, RM conditions yielded greater accuracy than the NO RM conditions. Of initial interest was the comparison of baseline conditions as shown in Table 4. The use of AP alone without the smartphone was compared to KEMAR alone to determine improvements related to the AP alone. Contrary to the results obtained with humans, the accuracy of the VTT was higher for the KEMAR alone condition compared to the two AP conditions, most likely due to the occlusion effect of the AP and lack of binaural processing. Therefore, given the uniqueness of this arrangement, no statistical analyses were completed for the baseline conditions.

Table 4. Mean transcription accuracy at each baseline condition.

Technology	Speech noise (dB SNR)			Babble noise (dB SNR)			Mean
	+5	0	-5	+5	0	-5	
AP NC	93.83	79.01	46.30	93.83	78.40	45.06	72.74
AP TP	85.80	67.28	45.06	92.59	78.40	59.26	71.40
KEMAR	96.00	79.67	49.00	93.67	83.67	60.67	77.11
Mean	91.88	75.32	46.79	93.36	80.16	55.00	73.75

NOTE: The mean transcription accuracy of three baseline testing conditions are presented as combinations of noise type and SNR: Babble/speech noise at +5, 0, and -5 dB SNR. AP = AirPods Pro, NC=noise cancellation, TP = transparency, SNR = signal-to-noise ratio, KEMAR = Knowles Electronics Manikin for Acoustic Research, KEMAR means the KEMAR alone condition.

Table 5. Mean transcription accuracy at each remote microphone condition.

Technology	Speech noise (dB SNR)			Babble noise (dB SNR)			Mean
	+5	0	-5	+5	0	-5	
AP NC + LL	98.67	95.33	86.33	95.67	92.00	87.00	92.50
AP TP + LL	94.33	84.33	72.00	91.67	79.00	66.00	81.22
Roger RM system	98.67	98.67	94.33	96.33	95.67	86.67	95.06
Mean	97.22	92.78	84.22	94.56	88.89	79.89	89.59

NOTE: The mean transcription accuracy of three RM testing conditions are presented as combinations of noise type and SNR: Babble/speech noise at +5, 0, and -5 dB SNR. AP = AirPods Pro, LL = Live Listen, NC = noise cancellation, TP = transparency, SNR = signal-to-noise ratio. .

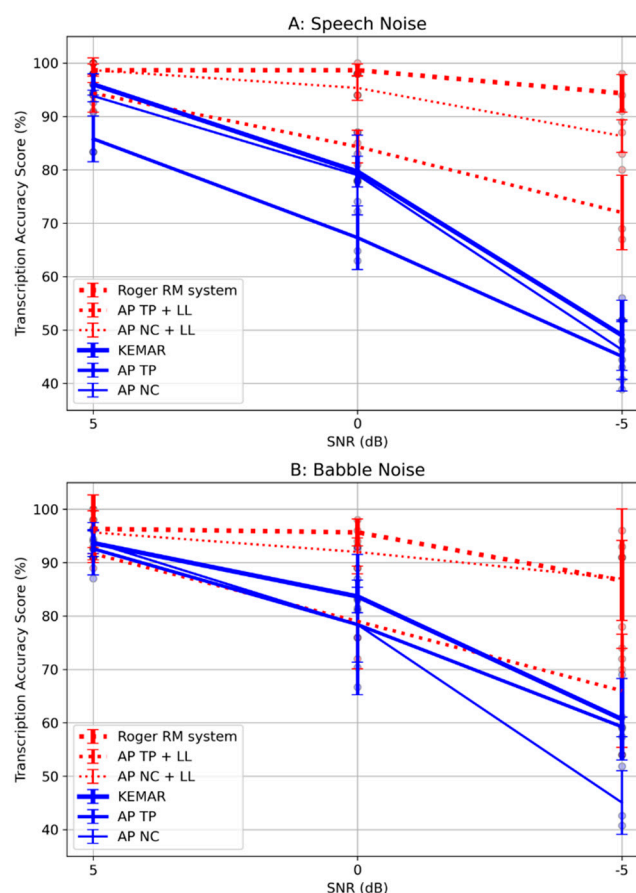


Figure 2. Mean and standard deviation of transcription accuracy scores under different conditions. A) under speech-shaped noise, B) under babble noise. NOTE: RM = remote microphone, AP = AirPods Pro, TP = transparency, NC = noise cancellation, LL = Live Listen, KEMAR = Knowles Electronics Manikin for Acoustic Research, SNR = signal-to-noise ratio.

Of greater interest was the comparison of the RM technology conditions shown in Table 5. The statistical analyses included a repeated-measures ANOVA for the RM technology conditions of interest (Roger, AP NC + LL and AP TP + LL), noise type (speech and babble), and SNR (+5, 0, and -5 dB). Significant main effects were observed for technology, $F(2,34) = 47.45$, $p < .001$, noise type, $F(1,34) = 8.65$, $p = .006$, and SNR, $F(2,34) = 42.91$, $p < .001$. The accuracy was significantly higher when tested in speech noise compared to babble noise (84.22% vs 79.89%). A significant two-way interaction was only found between SNR and technology, $F(4,34) = 6.03$, $p = .001$. All other interactions were non-significant ($p > .05$).

Follow-up analyses were performed using Tukey's adjustment for each significant effect (p -value was marked as p_{adj}). For the main effect of RM technology conditions, the transcription accuracy of Roger (95.06%) was significantly better than AP TP + LL (81.22%) ($t_{34} = 9.15$, $p_{adj} < .001$; large Cohen's $d = 1.50$) but not better than AP NC + LL (92.50%) ($t_{34} = 1.69$, $p_{adj} = .22$). There was also a significant difference between the two AP conditions that AP NC + LL was significantly greater than AP TP + LL ($t_{34} = 7.46$, $p_{adj} < .001$; large Cohen's $d = 1.17$). For the main effect of three SNR conditions, all comparisons (+5 vs. 0, 0 vs. -5, and +5 vs. -5 dB) were significant as expected, with $t_{34} = 3.35, 5.81, 9.15$, respectively, and all $p_{adj} < .002$. The relevant effect sizes were all large (Cohen's $d = .80, .85, 1.52$, respectively).

Figure 3 illustrates the two-way interaction between the settings of SNR and RM technology conditions across both noise types. The follow-up pairwise comparisons are shown in Table 6. The differences among the conditions were most evident at the more challenging SNR conditions (0 and

-5 dB). The patterns of significance at these SNRs were the same as the main effects, with Roger and AP NC + LL significantly greater than AP TP + LL but no significant difference between them.

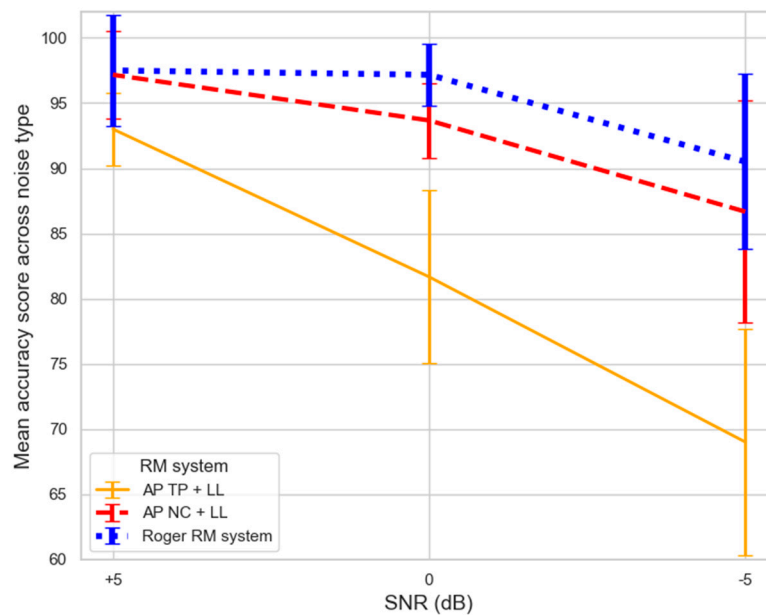


Figure 3. Averaged scores across noise types for the three RM conditions. NOTE: RM = remote microphone, AP = AirPods Pro, NC = noise cancellation, TP = transparency, SNR = signal-to-noise ratio, LL = Live Listen.

Table 6. Post-hoc analysis of pair-wise comparisons of remote microphone conditions under different signal-to-noise ratios.

SNR (dB)	+5	0	-5
Roger RM system vs. AP NC+LL	0.13	1.34	1.46
Roger RM system vs. AP TP+LL	1.72	5.92***	8.21***
AP NC+LL vs. AP TP+LL	1.59	4.58**	6.75***

NOTE: t value was indicated. Strength of Tukey adjusted p -values: * $p_{adj} < 0.05$, ** $p_{adj} < .01$, *** $p_{adj} < .001$. Same abbreviations as Table 5.

4. Discussion

The aim of the study was to compare the transcription accuracy in noise when using AP as part of an RM system (iPhone set to LL) and when using a sophisticated RM system (Phonak Roger). Three factors were involved including technology conditions, noise type, and SNR. The results indicated that the transcription accuracy was significantly influenced by all three factors, as well as a two-way interaction of SNR by technology. Overall, accuracy scores were lowest at -5 compared to 0 and +5 dB SNR, and highest when using an RM such as Roger On transmitting to a Roger receiver or using a smartphone with LL transmitting via Bluetooth low energy to an AP set to NC mode.

Regarding the baseline conditions (Table 4), it is interesting to note that using the AP alone didn't increase the accuracy score. Because of the somewhat artificial nature of the VTT arrangement with a manikin, the baseline results are provided for information purposes. As mentioned earlier, the results may be impacted by the fitting of the AP on an artificial pinna. It should also be noted that in this test arrangement using KEMAR, the result is obtained monaurally so that the benefits of binaural listening that are available with human listeners were not observed.

However, when using an RM system (Table 5), there were improvements in the transcription accuracy relative to using KEMAR alone. Valderrama et al. [22] reported that using AP on TH adults with self-disclosed difficulties hearing in noise, set for "maximum ambient noise reduction" and

“conversation boost” enabled, provided a 5.4 dB SNR advantage and 11.8% intelligibility increase. It is likely that if they had included use of the smartphone as an RM system, the improvements would have been greater. In addition, there was an 8% reduction in mental demand and listening effort. In the present study, the most accurate VTT score with AP was obtained when using LL with and AP also set to NC mode (92.5%). This agrees with Foroogzar [24] who reported 94.4% accurate sentence recognition on average when testing adults with typical hearing wearing AP with iPhone set to LL. The NC mode deactivates the microphones on the AP and allows for signals with the highest quality as the input to the Otter transcription program. Further research is required to confirm such benefits in humans with different degrees of hearing impairment.

When considering the interaction between SNR and technology, the most demanding listening conditions (0 and -5 dB SNR) revealed significant differences in accuracy among the three RM conditions. The accuracy was significantly lower when using AP TP + LL compared to Roger and AP NC + LL, although there was no significant difference between Roger and AP NC + LL. This suggests that RM features offer potential benefits relative to AP TP + LL mode which allows the transmission of environmental noise. In the AP TP mode, the environmental noise is mixed with the signal arriving from the RM thus reducing its potential benefit. When comparing smartphone-based RM to Roger systems, differences may be attributed to Bluetooth transmission delays [7] that are not present in the Roger RM system that uses digital modulation rather than Bluetooth.

These results have potential clinical and research implications. Using Roger and smartphone-based RM is beneficial in reducing the challenges caused by noise relative to not using any device across SNR and noise types, with benefits of 17.95, 15.39 and 4.11% in Roger, AP NC + LL and AP TP + LL, respectively. Considering the current device setting is mainly for those without hearing loss, such benefits can potentially help improve hearing difficulties in noise for listeners with TH, occupying 12-15% of the general population [1,2]. The benefits of the Roger RM system were confirmed by Thibodeau et al. [9] in 10 subjects with TH despite using different Roger transmitters (Pen, Select) and receivers (Roger Focus-first generation). Similar to the RM system, the benefits of using smartphone-based RM on speech recognition in noise are expected in humans with TH. It's also of possible benefit to those who have hearing aids that connect to smartphones when using a remote mic app on the phone if the Bluetooth delay can be tolerated. However, many HA manufacturers do offer proprietary RM devices now with personal hearing technology.

In addition, iPhone and AP can work as a portable RM system in daily communication for listeners with TH. This is similar to the findings in Table 1 on the benefit of using AP as a hearing assistance device [19–24]. Given the widespread adoption and convenience of smartphones and earbuds, along with the lower cost compared to Roger devices, smartphone-based RM systems can be considered as alternative RM systems for TH individuals with hearing difficulties in noise, particularly those concerned about cost or the stigma associated with traditional hearing devices. Additionally, with the governmental approval of the software in iPhone to adjust AP for persons with mild-to-moderate hearing impairments, such systems may function both as RM systems and as hearing assistive devices. However, further research on humans is required before these solutions can be suggested as part of clinical protocols.

Using VTT and KEMAR has the potential to provide objective verification of new technological features, especially in situations where real participants are unavailable due to constraints such as COVID-19 or lack of funding. This approach can yield objective results that are not influenced by human factors such as age, gender, emotional status, personality, cognitive functions, etc. However, to ensure comparability across different studies, the version of the transcription application (e.g., Otter) should be specified, given the rapid pace of feature development in this high-tech era.

Although the results suggested a promising testing method for research and highlighted the potential benefits of using RM for listeners with TH, there are limitations to consider. Firstly, the use of KEMAR may restrict the applicability of the results to humans. Further research involving carefully controlled variables (e.g., age, cognition, and hearing status) is necessary to validate these findings before adopting such a smartphone-based RM system for rehabilitative solutions. Secondly,

only one smartphone-based RM system using an iOS device and a single type of headphone was evaluated. This limitation restricts the generalizability of the results to other smartphone platforms (e.g., Android) and different headphones with similar features like LL. Finally, while the cost of an iPhone and AP may be lower than that of a sophisticated Roger system, these devices may still be expensive for individuals especially if they are not iPhone users.

5. Conclusions

The use of a manikin and a VTT paradigm showed that speech recognition in noise averaged across two RM systems was better in speech versus babble noise and improved as the SNR changed from -5 to +5 dB. These findings supported the validity of the methodology for comparing three RM conditions. In the more challenging noise conditions, the arrangement with the manikin receiving speech through the AP in NC mode and the smartphone set to LL program yielded comparable VTT accuracy to that obtained with the Roger On transmitter/Roger Focus receiver RM system. However, the considerable Bluetooth transmission delay when using the smartphone RM system may limit the acceptance of the AP/iPhone arrangement as a solution for improving speech recognition in noisy environments for persons with TH.

Author Contributions: Conceptualization, L.T.; methodology, L.T.; formal analysis, S.Q.; investigation, S.Q.; data curation, S.Q.; writing—original draft preparation, S.Q.; writing—review and editing, S.Q., and L.T.; visualization, S.Q.; supervision, L.T.; project administration, L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw VTT data reported in this study are available in Appendix: Table A1. Raw transcription accuracy data.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AP	AirPods Pro
AP 2	AirPods Pro 2
BKB	Bamford-Kowal-Bench
CB	Conversation Boost
HA	Hearing aid
HINT	Hearing in noise test
KEMAR	Knowles Electronics Manikin for Acoustic Research
LL	Live listen
MHINT	Mandarin hearing in noise test
NAL-NL2	National Acoustics Laboratories-Non-Linear prescription, 2nd generation
NC	Noise cancellation
OTC	Over-the-counter
PSAP	Personal sound amplification products
RM	Remote microphone
SNR	Signal-to-noise ratio
SRT	speech reception threshold
TP	Transparency
TH	Typical hearing
VTT	Voice-to-text

Appendix A

Appendix A.1

Table A1. Raw transcription accuracy data.

Noise	Condition	SNR	1st	2nd	3rd	Mean	SD
Speech	Roger	+5	98.00	98.00	100.00	98.67	1.15
Speech	Roger	0	98.00	98.00	100.00	98.67	1.15
Speech	Roger	-5	98.00	91.00	94.00	94.33	3.51
Speech	AP TP + LL	+5	91.00	98.00	94.00	94.33	3.51
Speech	AP TP + LL	0	81.00	87.00	85.00	84.33	3.06
Speech	AP TP + LL	-5	69.00	80.00	67.00	72.00	7.00
Speech	AP NC + LL	+5	96.00	100.00	100.00	98.67	2.31
Speech	AP NC + LL	0	94.00	98.00	94.00	95.33	2.31
Speech	AP NC + LL	-5	83.00	87.00	89.00	86.33	3.06
Speech	KEMAR	+5	96.00	94.00	98.00	96.00	2.00
Speech	KEMAR	0	78.00	78.00	83.00	79.67	2.89
Speech	KEMAR	-5	43.00	56.00	48.00	49.00	6.56
Speech	AP TP	+5	90.74	83.33	83.33	85.80	4.28
Speech	AP TP	0	62.96	64.81	74.07	67.28	5.95
Speech	AP TP	-5	44.44	51.85	38.89	45.06	6.50
Speech	AP NC	+5	94.44	94.44	92.59	93.83	1.07
Speech	AP NC	0	87.04	72.22	77.78	79.01	7.48
Speech	AP NC	-5	51.85	40.74	46.30	46.30	5.56
Babble	Roger	+5	89.00	100.00	100.00	96.33	6.35
Babble	Roger	0	93.00	98.00	96.00	95.67	2.52
Babble	Roger	-5	78.00	91.00	91.00	86.67	7.51
Babble	AP TP + LL	+5	91.00	93.00	91.00	91.67	1.15
Babble	AP TP + LL	0	72.00	89.00	76.00	79.00	8.89
Babble	AP TP + LL	-5	54.00	70.00	74.00	66.00	10.58
Babble	AP NC + LL	+5	91.00	98.00	98.00	95.67	4.04
Babble	AP NC + LL	0	89.00	93.00	94.00	92.00	2.65
Babble	AP NC + LL	-5	72.00	93.00	96.00	87.00	13.08
Babble	KEMAR	+5	94.00	91.00	96.00	93.67	2.52
Babble	KEMAR	0	81.00	83.00	87.00	83.67	3.06
Babble	KEMAR	-5	54.00	59.00	69.00	60.67	7.64
Babble	AP TP	+5	94.44	96.30	87.04	92.59	4.90
Babble	AP TP	0	92.59	66.67	75.93	78.40	13.14
Babble	AP TP	-5	57.41	59.26	61.11	59.26	1.85
Babble	AP NC	+5	92.59	92.59	96.30	93.83	2.14
Babble	AP NC	0	70.37	81.48	83.33	78.40	7.01
Babble	AP NC	-5	40.74	51.85	42.59	45.06	5.95

NOTE: AP = AirPods Pro, NC = noise cancellation, TP = transparency, SNR = signal-to-noise ratio, KEMAR means the KEMAR alone condition, Roger means the Roger remote microphone condition.

References

1. Tremblay, K. L.; Pinto, A.; Fischer, M. E.; Klein, B. E. K.; Klein, R.; Levy, S.; Tweed, T. S.; Cruickshanks, K. J. Self-Reported Hearing Difficulties among Adults with Normal Audiograms: The Beaver Dam Offspring Study. *Ear Hear* **2015**, *36* (6), e290-299. <https://doi.org/10.1097/AUD.0000000000000195>.
2. Spankovich, C.; Gonzalez, V. B.; Su, D.; Bishop, C. E. Self Reported Hearing Difficulty, Tinnitus, and Normal Audiometric Thresholds, the National Health and Nutrition Examination Survey 1999–2002. *Hearing Research* **2018**, *358*, 30–36. <https://doi.org/10.1016/j.heares.2017.12.001>.
3. *Practice guidelines and standards*. American Academy of Audiology. <https://www.audiology.org/practice-resources/practice-guidelines-and-standards/> (accessed 2025-03-16).
4. *ASHA practice policy*. American Speech-Language-Hearing Association. <https://www.asha.org/policy/> (accessed 2025-03-16).
5. *Position statement: Preventing noise-induced occupational hearing loss*. American Academy of Audiology. <https://www.audiology.org/practice-guideline/position-statement-preventing-noise-induced-occupational-hearing-loss/> (accessed 2025-03-16).
6. Mealings, K.; Yeend, I.; Valderrama, J. T.; Gilliver, M.; Pang, J.; Heeris, J.; Jackson, P. Discovering the Unmet Needs of People with Difficulties Understanding Speech in Noise and a Normal or Near-Normal Audiogram. *American Journal of Audiology* **2020**, *29* (3), 329–355. https://doi.org/10.1044/2020_AJA-19-00093.
7. Thibodeau, L. M. Between the Listener and the Talker: Connectivity Options. *Semin Hear* **2020**, *41* (4), 247–253. <https://doi.org/10.1055/s-0040-1718710>.
8. Thibodeau, L. M. Benefits in Speech Recognition in Noise with Remote Wireless Microphones in Group Settings. *J Am Acad Audiol* **2020**, *31* (6), 404–411. <https://doi.org/10.3766/jaaa.19060>.
9. Thibodeau, L. M.; Land, V.; Sivaswami, A.; Qi, S. Benefits of Speech Recognition in Noise Using Remote Microphones for People with Typical Hearing. *Journal of Communication Disorders* **2024**, 106467. <https://doi.org/10.1016/j.jcomdis.2024.106467>.
10. Shiels, L.; Tomlin, D.; Rance, G. The Assistive Benefits of Remote Microphone Technology for Normal Hearing Children with Listening Difficulties. *Ear and Hearing* **2023**, *44* (5), 1049. <https://doi.org/10.1097/AUD.0000000000001351>.
11. Mealings, K.; Valderrama, J. T.; Mejia, J.; Yeend, I.; Beach, E. F.; Edwards, B. Hearing Aids Reduce Self-Perceived Difficulties in Noise for Listeners with Normal Audiograms. *Ear Hear* **2024**, *45* (1), 151–163. <https://doi.org/10.1097/AUD.0000000000001412>.
12. da Silva, J. C.; de Araujo, C. M.; Lüders, D.; Santos, R. S.; Moreira de Lacerda, A. B.; José, M. R.; Guarinello, A. C. The Self-Stigma of Hearing Loss in Adults and Older Adults: A Systematic Review. *Ear Hearing* **2023**, *44* (6), 1301. <https://doi.org/10.1097/AUD.0000000000001398>.
13. Shankar, N.; Bhat, G. S.; Panahi, I. M. S.; Tittle, S.; Thibodeau, L. M. Smartphone-Based Single-Channel Speech Enhancement Application for Hearing Aids. *The Journal of the Acoustical Society of America* **2021**, *150* (3), 1663–1673. <https://doi.org/10.1121/10.0006045>.
14. *Mobile fact sheet*. Pew Research Center: Internet, Science & Tech. <https://www.pewresearch.org/internet/fact-sheet/mobile/> (accessed 2025-03-16).
15. Thibodeau, L. M.; Panahi, I. Smartphone Technology: The Hub of a Basic Wireless Accessibility Plan. *The Hearing Journal* **2021**, *74* (4), 6. <https://doi.org/10.1097/01.HJ.0000743696.05586.db>.
16. U.S. Food and Drug Administration. *FDA authorizes first over-the-counter hearing aid software*. <https://www.fda.gov/news-events/press-announcements/fda-authorizes-first-over-counter-hearing-aid-software> (accessed 2025-03-16).
17. Gilmore, J. N. Design for Everyone: Apple AirPods and the Mediation of Accessibility. *Critical Studies in Media Communication* **2019**, *36* (5), 482–494. <https://doi.org/10.1080/15295036.2019.1658885>.
18. Nelson, J.; Dunn, A. *Assistive listening device for children - roger focus II*. Phonak. https://www.phonak.com/content/dam/phonak/en/evidence-library/field-studies/PH_FSN_Roger_Focus_II_in_children_with_%20UHL_210x280_EN_V1.00.pdf (accessed 2025-03-16).
19. Lin, H.-Y. H.; Lai, H.-S.; Huang, C.-Y.; Chen, C.-H.; Wu, S.-L.; Chu, Y.-C.; Chen, Y.-F.; Lai, Y.-H.; Cheng, Y.-F. Smartphone-Bundled Earphones as Personal Sound Amplification Products in Adults with Sensorineural Hearing Loss. *iScience* **2022**, *25* (12), 105436. <https://doi.org/10.1016/j.isci.2022.105436>.

20. Hammond, E.; Diedesch, A. C. Usability and Perceived Benefit of Hearing Assistive Features on Apple AirPods Pro. *Proceedings of Meetings on Acoustics* **2023**, *51* (1), 050004. <https://doi.org/10.1121/2.0001771>.
21. Martinez, S. *Evaluating apple AirPods pro 2 hearing protection and listening*. The Hearing Review. <https://hearingreview.com/inside-hearing/research/evaluating-apple-airpods-pro-2-for-hearing-protection-and-listening> (accessed 2025-03-16).
22. Valderrama, J. T.; Mejia, J.; Wong, A.; Chong-White, N.; Edwards, B. The Value of Headphone Accommodations in Apple AirPods pro for Managing Speech-in-Noise Hearing Difficulties of Individuals with Normal Audiograms. *International Journal of Audiology* **2024**, *63* (6), 447–457. <https://doi.org/10.1080/14992027.2023.2199442>.
23. Kim, G.-Y.; Yun, H. J.; Jo, M.; Jo, S.; Cho, Y. S.; Moon, I. J. Apple AirPods pro as a Hearing Assistive Device in Patients with Mild to Moderate Hearing Loss. *Yonsei Medical Journal* **2024**, *65*. <https://doi.org/10.3349/ymj.2023.0375>.
24. Foroogozar, M. *Assessing the Influence of Apple AirPods with Live Listen Feature on Speech Recognition and Memory Retention in Noise Levels Simulating Noisy Healthcare Settings - Insights from QuickSIN*; Arizona State University, 2024. <https://keep.lib.asu.edu/items/193345> (accessed 2025-03-16).
25. Goldenthal, E.; Park, J.; Liu, S. X.; Mieczkowski, H.; Hancock, J. T. Not All AI Are Equal: Exploring the Accessibility of AI-Mediated Communication Technology. *Computers in Human Behavior* **2021**, *125*, 106975. <https://doi.org/10.1016/j.chb.2021.106975>.
26. Millett, P. Accuracy of Speech-to-Text Captioning for Students Who Are Deaf or Hard of Hearing. *Journal of Educational, Pediatric & (Re)Habilitative Audiology* **2021**, *25*, 1–13.
27. Nilsson, M.; Soli, S. D.; Sullivan, J. A. Development of the Hearing in Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise. *The Journal of the Acoustical Society of America* **1994**, *95* (2), 1085–1099. <https://doi.org/10.1121/1.408469>.
28. Zanin, J.; Rance, G. Functional Hearing in the Classroom: Assistive Listening Devices for Students with Hearing Impairment in a Mainstream School Setting. *International Journal of Audiology* **2016**, *55* (12), 723–729. <https://doi.org/10.1080/14992027.2016.1225991>.
29. Van Engen, K. J. Speech-in-Speech Recognition: A Training Study. *Language and Cognitive Processes* **2012**, *27* (7–8), 1089–1107. <https://doi.org/10.1080/01690965.2012.654644>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.