# WSCoach: Wearable Real-time Auditory Feedback for Reducing Unwanted Words in Daily Communication

ZHANG YOUPENG, School of Creative Media, City University of Hong Kong, China
NUWAN JANAKA, Smart Systems Institute, National University of Singapore, Singapore
ASHWIN RAM, Saarland Informatics Campus, Saarland University, Germany and School of Creative Media, City University of Hong Kong, China
YIN PEILIN, Zhejiang University, China
TIAN YANG, School of computer, electronics, and information, Guangxi University, China
SHENGDONG ZHAO*, School of Creative Media & Department of Computer Science, City University of Hong Kong, China
PIERRE DRAGICEVIC*, Inria, CNRS, Université de Bordeaux, France

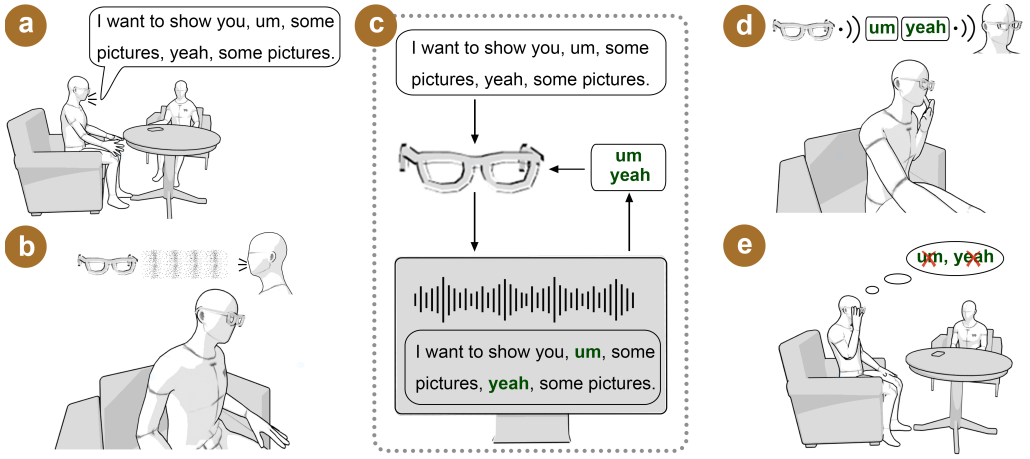


Fig. 1. *WSCoach* helps users reduce the use of words they wish not to say in daily conversations, such as filler words and swear words. (a) The user wears smart audio-based glasses and engages in daily conversations; (b) The (audio) smart glasses continuously monitors their speech, which is (c) analyzed to identify the unwanted words the user has specified in advance; (d) *WSCoach* provides real-time auditory feedback through the smart glasses. (e) This active self-monitoring helps the user refrain from using unwanted words during subsequent interactions. The above steps also represent the experimental task in Section 6.

*Corresponding Authors.

Authors' Contact Information: Zhang Youpeng, yzhan63@cityu.edu.hk, School of Creative Media, City University of Hong Kong, Hong Kong, China; Nuwan Janaka, nuwanj@u.nus.edu, Smart Systems Institute, National University of Singapore, Singapore; Ashwin Ram, ashwinram10@gmail.com, Saarland Informatics Campus, Saarland University, Saarland, Germany, Synteraction Lab and School of Creative Media, City University of Hong Kong, Hong Kong, China; Yin Peilin, 3200105215@zju.edu.cn, Zhejiang University, Hangzhou, China; Tian Yang, ytian@gxu.edu.cn, School of computer, electronics, and information, Guangxi University, Nanning, China; Shengdong Zhao, shengdong.zhao@cityu.edu.hk, School of Creative Media & Department of Computer Science, City University of Hong Kong, Hong Kong, China; Pierre Dragicevic, pierre.dragicevic@inria.fr, Inria, CNRS, Université de Bordeaux, Bordeaux, France.

The rise of wearable smart devices raises unprecedented opportunities for self-improvement through ubiquitous behavior tracking and guidance. However, the design of effective wearable behavior intervention systems remains relatively unexplored. To address this gap, we conducted controlled studies focusing on the reduction of unwanted words (e.g., filler words, swear words) in daily communication through auditory feedback using wearable technology. We started with a design space exploration, considering various factors such as the type, duration, and timing of the auditory feedback. Then, we conducted pilot studies to reduce the space of design choices and prototyped a system called *WSCoach* (Wearable Speech Coach), which informs users when they utter unwanted words in near-real-time. To evaluate *WSCoach*, we compared it with a state-of-the-art mobile application supporting post-hoc conversation analysis. Both approaches were effective in reducing the occurrence of unwanted words, but *WSCoach* appears to be more effective in the long run. Finally, we discuss guidelines for the design of wearable audio-based behavior monitoring and intervention systems and highlight the potential of wearable technology for facilitating behavior correction and improvement. For supplementary material, please see the META Appendix and our OSF project at https://osf.io/6vhwn/?view_only=489498d3ac2d4703a17475fc6ca65dfa.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**; **Empirical studies in ubiquitous and mobile computing**; **Mobile devices**.

Additional Key Words and Phrases: audio feedback, notification, filler words, speech coach, spearcon, habit reversal, awareness training, conversation

**ACM Reference Format:**
Zhang Youpeng, Nuwan Janaka, Ashwin Ram, Yin Peilin, Tian Yang, Shengdong Zhao, and Pierre Dragicevic. 2025. WSCoach: Wearable Real-time Auditory Feedback for Reducing Unwanted Words in Daily Communication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 152 (September 2025), 30 pages. https://doi.org/10.1145/3749531

## 1 Introduction

The rise of smart wearable devices has opened up unprecedented opportunities for improving our daily lives through ubiquitous behavior tracking and guidance. One promising area of research involves utilizing wearable devices to monitor users' status and deliver intelligent interventions to improve well-being and behaviors [11, 71]. While visual feedback-focused wearable solutions have demonstrated effectiveness (e.g., [20, 74, 81]), it is essential to recognize that our visual channels are frequently occupied [8]. Consequently, audio-based real-time monitoring and intervention solutions hold considerable promise [8, 46] despite being relatively less explored.

In particular, we aim to leverage audio-based solutions to tackle speech-related behaviors. Effective communication is vital in both personal and professional contexts [72, 86], allowing us to navigate diversity, foster trust and respect, and cultivate an environment conducive to sharing ideas and problem-solving. Nevertheless, it is uncommon to find individuals who can entirely refrain from using filler words or repetitive expressions, especially when hurried or unprepared in speech. These unwanted words, encompassing filler words, unprofessional slang, offensive language and repetitive expressions, are widely recognized as superfluous language that can undermine the effectiveness of communication, impacting the speaker's credibility and the audience's comprehension [72].

Unlike activities like running or cycling, where existing audio-based intervention solutions have been designed and effectively used (e.g., [6]), everyday conversations involve social situations where the potential interference [50] of the feedback provided by the wearable solution to one's social acceptance needs to be carefully considered. Disruptive intervention techniques can be seen as impolite and undesirable [1], thus demanding a separate investigation. While existing interventions leverage mobile applications to assist individuals in curbing the usage


ACM 2474-9567/2025/9-ART152
https://doi.org/10.1145/3749531

of unwanted words (e.g., [61]), these solutions may disrupt natural conversations and prove inconvenient in hands-free scenarios.

To address this challenge, we introduce Wearable Speech Coach (*WSCoach*), a system designed to enhance awareness through carefully designed auditory feedback, providing near real-time interventions (with feedback provided within 1 to 2 seconds) to decrease the usage of unwanted words. *WSCoach* can be adapted to any wearable device with audio input and output, such as Bluetooth earphones or headsets. While wrist-worn devices and smart rings offer discreetness, they lack the ability to deliver detailed, real-time auditory feedback during conversation. Smart glasses with visual displays support richer feedback but can disrupt eye contact and visual attention—both critical for face-to-face interaction. To balance functionality and social appropriateness, we selected an audio-based smart glasses platform. This form factor enables subtle, always-available feedback with minimum interference to conversational norms. Unlike mobile phones or earphones, which may appear distracting or impolite, audio-based smart glasses provide a more seamless and socially acceptable alternative. The feedback is mostly audible only to the wearer, minimizing disruption to others and supporting natural, real-time interaction. Additionally, their hands-free, heads-up design supports mobility, multitasking, and privacy—aligning with our goal of delivering discreet, context-sensitive support. These qualities also reflect the heads-up computing paradigm [87], which seeks to embed computational support seamlessly into everyday social interactions. While smart glasses remain an emerging platform, their expanding ecosystem and growing market presence suggest strong potential for broader adoption.

We conducted a series of pilot studies to identify the optimal attributes of auditory feedback (i.e., type, duration, timing), ultimately revealing that spearcon emerged as the most preferred feedback mechanism based on consistently higher positive ratings across various metrics. Building on these findings, we evaluated the efficacy of *WSCoach* in reducing the occurrence of unwanted words during conversations, comparing its performance to the professional mobile application, "*Orai*". The results confirmed the effectiveness of *WSCoach* in decreasing unwanted words.

The contributions of this paper are thus threefold: 1) Understand and empirically evaluate the suitability of various auditory feedback to improve awareness of speaking unwanted words during daily communications. 2) An artifact that helps people reduce unwanted words using wearable technology with real-time auditory feedback. 3) An empirical study that evaluates the effectiveness of decreasing unwanted words with a wearable system, offering further design implications.

## 2 Related Work

Speech disfluencies, such as filled pauses (e.g., "uh", "um"), tongue clicks, and frequently used fillers (e.g., "like", "you know"), are prevalent in everyday communication and can hinder effective communication [48]. People are motivated to minimize such behaviors (e.g., unwanted words) and improve their daily communication [53]. Recent advances in wearable technologies have opened new avenues for interventions aimed at enhancing communication [9, 20, 56]. Our studies focus on reducing unwanted words in conversations using (near) real-time speech intervention with auditory feedback, contributing to the broader objective of facilitating real-time speech improvement through wearable devices. Thus, our research is related to:

### 2.1 Speech Interventions and Awareness Training

Speech interventions aim to enhance communication proficiency and reduce speech disfluencies [31, 55, 62, 75]. While there are different types of speech interventions (see [53] for a review), awareness training is a common and effective strategy to reduce speech disfluencies [32, 53, 72]. Awareness training involves the conscious identification of specific undesirable speech patterns, such as filled pauses or the overuse of filler phrases like "like" and "you know," facilitated by various means, including audio and video recordings [53].

Awareness training ranges from (near) real-time methods[1], where a trainer signals the occurrence of disfluencies (e.g., by alerting feedback) [31, 73, 75], to post-training analysis (e.g., retrospective feedback) through audio/video recordings to pinpoint disfluencies [13, 62]. Early real-time methods suffered from a generic feedback approach, which lacked specificity for different types of disfluencies. For example, these early works used a human observer monitoring the user's speech behaviors and raising hands [55, 75] or initiating a single auditory alert [31, 73] upon detecting any speech disfluency. While this approach enabled users to identify the disfluencies, it prevented users from distinguishing between them to have more detailed awareness (e.g., separating the filler word "like" from "you know").

Post-training techniques mitigate this lack of specificity by analyzing recorded speech to detect disfluencies and offering targeted suggestions as retrospective feedback, but they demand additional time and effort for post-analysis. Most of these post-analysis methods use desktop or mobile platforms to provide detailed feedback due to their processing and feedback capabilities. For example, VoiceCoach [84] is an interactive desktop application targeting public speaking that analyzes voice modulation skills, recommends suitable examples (e.g., from TED talks), and provides visual feedback on performance (e.g., pause, volume, pitch, speed). Orai [61] is an interactive mobile application that provides retrospective feedback on speech components such as filler words, pace, and conciseness. It outputs a detailed analysis of recorded speeches via annotated transcripts, highlights areas of improvement in the user's speech, and helps keep track of progress. ELSA Speak [16] improves English speaking, specifically pronunciation, by utilizing AI to identify errors and provide visual feedback on recorded audio using a mobile application. StammerApp [51] is a mobile application designed to support people who stammer. It allows users to set goals related to challenging real-world speaking situations with recorded video/audio (e.g., ordering food, booking a taxi) and practice to overcome side effects (e.g., word repetition, regular use of interjections) during those situations .

Considering the trade-offs in previous systems—early (near) real-time feedback lacking specificity and post-intervention methods having specificity but requiring time or motivation to reflect—we consider the potential of (near) real-time feedback to support awareness training while maintaining specificity to minimize speech disfluencies.

## 2.2 Wearable Real-Time Interventions and Feedback Modalities

Wearable technology enables us to receive suitable real-time assistance (e.g., reminders from smartwatches, language translations from earbuds, navigation guidance from smart glasses) anywhere with little user effort. Wearable (near) real-time interventions have been commonly used in awareness training, including speech training [9, 20, 54, 81].

Most of these interventions use visual feedback due to its increased modality capacity [69] to show specific details. For example, Rhema [81] uses smart glasses to provide real-time visual feedback on the speaker's volume (e.g., make louder) and speech rate (e.g., slow down) during public speaking. Logue [20] analyzes verbal and non-verbal behavior during public speaking using sensors and provides visual feedback on behavior (speech rate, energy, openness, appropriateness) on smart glasses in an unobtrusive way. However, in conversations unlike public speaking, such visual feedback can disrupt social interactions due to reduced eye contact when attending to visual feedback [36, 50].

Thus, alternative feedback modalities like haptic feedback have been explored. AwareMe [9] is a wristband that provides real-time haptic feedback for anxiety detection during presentations. While haptic feedback is subtle and less interfering [79], it has limited modality capacity to show detailed feedback [9, 69]. As a result, AwareMe [9] also used visual feedback to indicate pitch, speed, and filler words.

[1]Note: There is a delay between actual occurrence and alerting feedback due to the detection time.

Thus, auditory feedback emerges as an alternative that strikes a balance between these two factors as it has higher modality capacity [69]. Therefore, real-time auditory feedback alone has been explored in various behavior interventions [41, 66, 82]. Radhakrishnan et al. [67] utilized wearables for real-time auditory reminders against unhealthy head postures, Casamassima et al. [12] for gait correction, and Md Mahbubur et al. [68] for guided breathing exercises.

### 2.3 Potential of Real-Time Auditory Feedback-Based Speech Interventions

In addition to the above-mentioned advantages, wearable auditory feedback systems can be more energy efficient than their visual feedback counterparts and thus have longer usage durations (e.g., auditory smart glasses generally last longer than visual smart glasses on a single charge [70]).

Existing (near) real-time auditory feedback systems in speech training typically target a general group of disfluencies (e.g., filler occurrences) [31] without providing detailed feedback (e.g., specific filler words). Given the higher modality capacity [69], (near) real-time auditory feedback can be used to indicate specific target behaviors during awareness training, which is underexplored in speech interventions. This underutilization may be due to the fact that auditory feedback during speech may disrupt natural speech flow by interfering with the speaking process itself (e.g., auditory masking[2]) [35, 64].

Furthermore, whether providing detailed auditory feedback in speech interventions can improve awareness and which types of auditory feedback are more suitable lacks systematic exploration in the literature. Thus, our work explores suitable (near) real-time auditory feedback for awareness training, utilizing wearable technology to develop and evaluate auditory feedback for speech correction, particularly in reducing unwanted words during conversations. Such interventions will highlight the potential of wearable intelligent assistants using auditory feedback and heads-up computing [80, 87, 88].

## 3 Literature Review: Auditory Feedback for Speech Interventions

To outline a design space of auditory feedback suitable for speech intervention, we reviewed papers from the fields of HCI and Speech Communication. Our review process focused on two guiding questions: 1) What constitutes auditory feedback interventions? and 2) What measures are suitable for evaluating the effectiveness of auditory feedback for speech interventions?

### 3.1 Methodology

To construct this design space, we employed a design space analysis method [23, 47]. Initially, we conducted searches in the ACM Digital Library, IEEE Xplore, Scopus, and Google Scholar using related keywords such as "*auditory feedback*", "*speech improvement*", "*speech intervention*", "*filler words*", "*self-awareness*", and "*communication aids*", following related systematic reviews [26, 53, 59, 65]. Our inclusion criteria comprised peer-reviewed publications aligned with our objectives, excluding works focused on non-auditory feedback or interventions, non-English publications, and publications before 2009. Subsequently, we reviewed the references of these relevant publications and included seminal works that introduced types of auditory feedback and auditory interventions, resulting in a total of 27 papers. Finally, we manually reviewed these papers and extracted themes that matched our objective (e.g., auditory feedback and its suitability for speech intervention).

### 3.2 Design Space

Based on our literature review, we have created a design space, as illustrated in Figure 2, which encompasses attributes of auditory feedback (e.g., how to feedback, what, when) and evaluation criteria for speech interventions.

[2]Masking refers to the reduction in sound clarity due to interference from another sound, such as the decreased comprehensibility of speech overlaid with background noise [64].

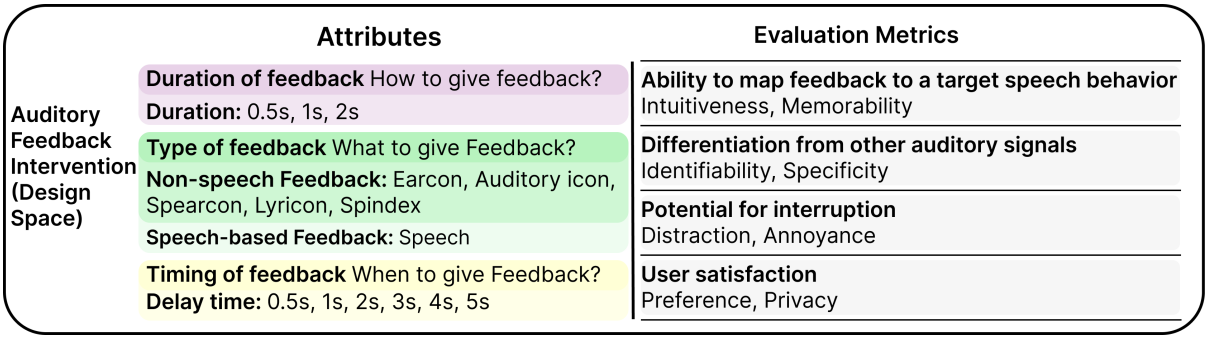


Fig. 2. The design space of auditory feedback for speech interventions.

*3.2.1 Auditory Feedback Intervention.* The design space of auditory feedback intervention comprises three attributes: how to give feedback, what to give as feedback, and when to give feedback. While we note that there are granular-level generic auditory feedback properties (e.g., frequency, loudness [10, 38, 64]) that can affect speech interventions, we do not consider them within our design space as they are determined when an auditory signal is designed (e.g., frequency/pitch) or are easily adjustable (e.g., loudness/amplitude).

*Duration of Feedback: How to give feedback?* Feedback duration is a critical factor that affects the user's experience during interventions [10]. For example, a prolonged sound can be distracting, especially if it interrupts ongoing tasks or occurs in a context where sustained attention is required. Conversely, a too-short sound can be difficult to identify, leading to a failure to respond appropriately to the feedback. While there are some guidelines on auditory feedback duration—such as more than 200 ms for vehicle warnings [38, 44], and between 400 ms and 2000 ms for auditory icons [10]—there is no consensus on the appropriate duration for speech interventions.

*Type of Feedback: What to give as Feedback?* Auditory feedback can be divided into two broad cue types: non-speech and speech-based feedback based on their common usage in digital interfaces [19, 38, 59, 64] (Please refer to the video figure to hear examples of each). Non-speech feedback includes earcons [5], which are non-speech audio cues, typically consisting of short, rhythmic sequences of musical notes with variable intensity, timbre, and register; auditory icons [10, 28, 57], which are nonmusical sounds that resemble (either literally or metaphorically) the objects, functions, and events they represent; spearcons [83], which are brief sounds produced by speeding up spoken phrases, sometimes to the point where the resulting sound is no longer comprehensible as speech; lyricons [37], which combine the lyrics (i.e., speech) with earcons; and spindexes [39], which are created by associating an auditory cue with an item (e.g., menu item), where the cue is based on the pronunciation of the first letter of each item. In contrast, speech-based feedback consists of spoken verbal cues or phrases. Given the advantages and disadvantages of cue type in various application scenarios [27, 59, 64]—such as auditory icons that rely on familiar, real-world sounds for intuitive understanding but might need learning to associate with specific actions; earcons that are flexible, abstract sounds suitable for hierarchical information but challenging to learn without inherent meaning; and spearcons that combine the clarity of speech with the flexibility of earcons and are less recognizable as speech, potentially reducing interference but may still suffer from speech-related issues like noise interference—the type of auditory feedback significantly influences its suitability for speech interventions.

*Timing of Feedback: When to give Feedback?* Auditory feedback can be used to indicate target behavior (e.g., unwanted words) during speech interventions (e.g., awareness training [53]) and reduce such behaviors [31, 48, 53, 55]. However, due to the temporal nature of the auditory modality and limitations of auditory sensory

memory (a.k.a., phonological short-term memory or echoic memory) [18, 21], such feedback can be provided either immediately or with a slight delay of up to 4 seconds to maintain in short-term memory for identification [21]. Moreover, the optimal timing for feedback may depend on the situation. For example, in high-stakes or safety-critical environments, immediate feedback might be crucial [44]. In contrast, in less critical learning situations, a brief delay might allow for reflection before receiving feedback [31]. Therefore, the 'delay time', the temporal gap between the target behavior (e.g., the utterance of unwanted words) and the corresponding auditory feedback, can influence the effectiveness of the auditory feedback intervention.

*Summary.* In auditory feedback for speech interventions, selecting between non-speech and speech-based cues requires careful consideration of the context-specific advantages. Feedback duration is a critical factor that necessitates a balance between impact and the potential for distraction. Timing is crucial, as the strategic delivery of feedback during speech hinges on the temporal nature of auditory perception and the memory constraints of users. While certain contexts provide established guidelines (e.g., vehicle warnings [38, 44]), due to the lack of such guidelines for speech interventions, identifying the optimal feedback attributes for speech interventions is a key area for further research. This underscores the intricate interplay between duration, type, and timing of auditory feedback in enhancing communication effectiveness.

*3.2.2 Evaluating Auditory Feedback Intervention.* Our literature review revealed a range of metrics for evaluating auditory feedback interventions across various applications [27, 37–39, 57, 64, 78, 83]. We compiled a list of 20 metrics to identify the most relevant metrics for speech interventions and engaged three co-authors in an independent review process. This collaborative effort led to the consensus selection of eight key metrics.

While considering these metrics, we assessed their applicability to our specific context. For example, the International Standards for Auditory Display Guidelines in Vehicles (ISO 15006) recommend metrics such as Audibility, Comprehensibility, Distinguishability, and Safety criticality [38, 77]. Given that our intervention scenarios primarily involve stationary settings rather than dynamic environments like driving, we deemed the Safety criticality metric less relevant and excluded it from our selection.

These metrics are essential for understanding the alignment between feedback objectives and user expectations. They include the ability to map feedback to a target speech behavior (e.g., Intuitiveness, Memorability), differentiation from other auditory signals (e.g., Identifiability, Specificity), the potential for interruption (e.g., Distraction, Annoyance), and user satisfaction (e.g., Preference, Privacy).

- **Intuitiveness** [10, 27, 38, 57] assesses the user's ability to understand the feedback's connection to the target speech behavior.
- **Memorability** [27, 58] evaluates how easily users can remember the meaning associated with specific auditory feedback.
- **Identifiability** [10, 38] determines whether the auditory feedback is easily distinguishable from other environmental sounds.
- **Specificity** [2, 38] measures the precision with which the feedback identifies the target speech behavior.
- **Distraction** [25, 56] examines the degree to which the feedback might interrupt the speech flow.
- **Annoyance** [39, 55] evaluates any discomfort or irritation caused by the feedback.
- **Preference** [10, 27, 39] investigates the types of feedback users find most agreeable and useful.
- **Privacy** [40, 63] considers the likelihood of the feedback inadvertently revealing sensitive information.

## 4 Pilot Studies: Designing Auditory Feedback to Minimize Unwanted Word Usage

To address our research question, *"What are the suitable attributes (i.e., duration, type, delay time) of auditory feedback for reminding users of unwanted word usage during daily conversations?"*, we conducted three iterative

pilot studies. Each study focused on one attribute. All user studies (the pilot studies here and the evaluation study we will report in Sec 6) have been approved by the Institutional Review Board at our university.

*4.0.1 Participants.* We recruited six participants (P1–P6, 4 male, 2 female, mean age = 23, SD = 1.4), all with no reported auditory impairments and self-identified as fluent English speakers, from the university community. All six participants did all three pilot studies, which took place on different days.

*4.0.2 Apparatus.* During conversations, participants wore smart audio glasses (Huawei Eyewear version 1[3] with lenses removed, 31g, stereo audio) connected to a mobile phone (Huawei P10) via Bluetooth to receive auditory feedback on unwanted word usage. The mobile phone was equipped with a custom-built Android application capable of triggering various types of auditory feedback, each with different durations and delay times.

## 4.1 Task and Procedure

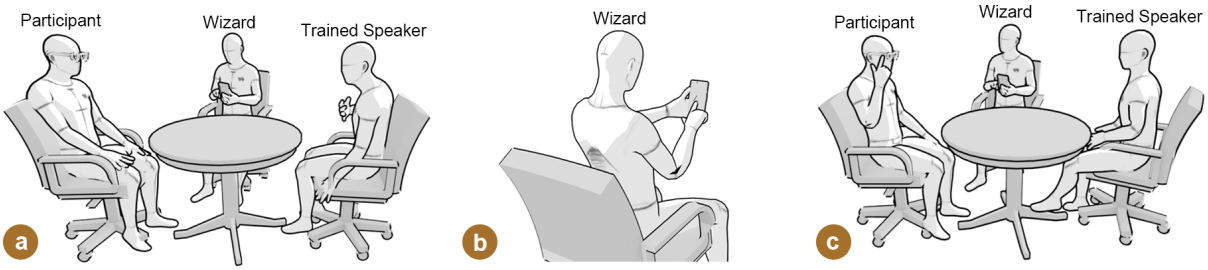


Fig. 3. Procedure for the pilot studies: (a) A participant, seated with two experimenters, engages in a conversation while wearing smart audio glasses. One experimenter (the trained speaker) converses with the participant, while the other (the wizard) operates the mobile app. (b) Upon detecting unwanted words spoken by the participant, the wizard activates auditory feedback via the app. (c) The participant receives auditory feedback through the smart audio glasses.

All pilot studies used Wizard of Oz testing (Figure 3). After obtaining consent and being provided with a briefing, each participant was asked to list unwanted words they frequently use in conversations and wished to minimize, encompassing both non-speech (e.g., "Um", "Uh") and speech (e.g., "I mean", "Like") words. Before the start of each pilot study, each participant engaged in a 10-minute training session encompassing every condition to get familiar with the feedback. During these sessions, they received auditory feedback designed for each condition, receiving feedback upon speaking the unwanted words they aimed to reduce. A trained speaker (acting as the conversation partner) facilitated the conversations, encouraging participants to speak on communication topics derived from IELTS Speaking Questions (https://ielts.org/). Another trained experimenter, referred to as the 'wizard,' monitored the conversation and activated predetermined auditory feedback whenever an unwanted word was detected. This auditory feedback was delivered to the participant through smart audio glasses, reminding them of their unwanted word usage, in line with awareness training practices [53]. Following each session, participants evaluated every auditory-feedback condition on the design-specific metrics (e.g., Distraction, Preference; see Sec. 3.2.2) using 7-point Likert scales.

[3]Web page from the manufacturer archived at https://web.archive.org/web/20221007024433/https://consumer.huawei.com/en/wearables/huawei-eyewear/. The Huawei Eyewear is crafted for continuous wear (6 hours of audio playback), featuring discreet/subtle acoustic outlets. The product has been replaced by a version 2.

## 4.2 Study Design

We conducted three pilot studies (Pilot 1: identifying suitable duration, Pilot 2: identifying suitable feedback type, and Pilot 3: identifying the acceptable feedback delay, in sequential order) to examine different **design attributes of auditory feedback**. Each study's findings informed the subsequent study, ensuring an iterative refinement process. The studies followed a within-subjects design: all participants (N=6) experienced all conditions. In Pilots 1 and 3, we tested three different duration conditions and six different timing conditions separately. Pilot 2 examined five types of auditory feedback. As a result, for Pilots 1 and 3, we counterbalanced the conditions using a Latin Square design, while in Pilot 2, with six participants and five conditions, a Latin Square design was infeasible, so we randomized the feedback types for each participant.

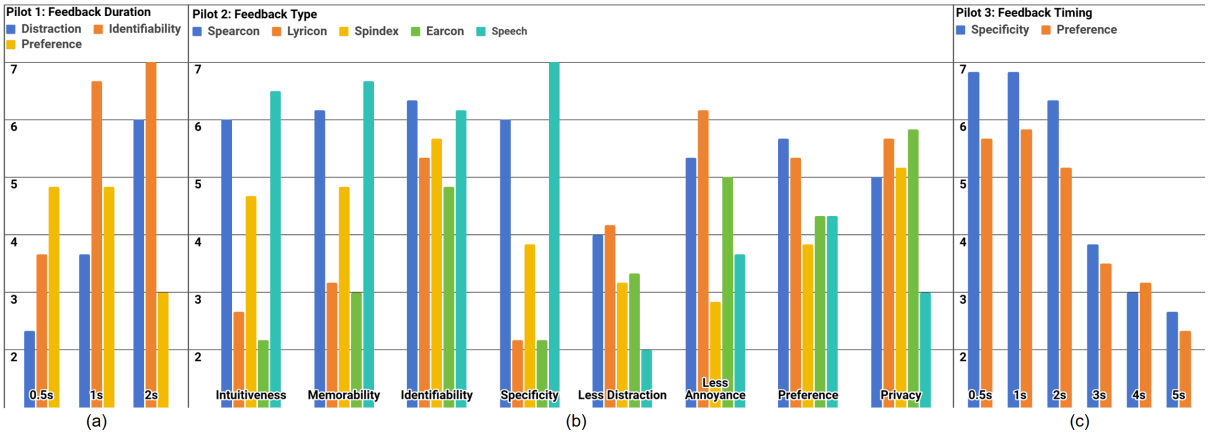


Fig. 4. Participants' (N=6) ratings for the (a) Pilot 1: Feedback Duration, (b) Pilot 2: Feedback Type, and (c) Pilot 3: Feedback Timing.

## 4.3 Pilot 1: Feedback Duration

*Design.* We tested three feedback durations: 0.5s, 1s, and 2s, focusing on measuring distraction, identifiability (i.e., if the feedback is easily distinguishable from other environmental sounds, see Sec 3.2), and user preference. Earcons were chosen for auditory feedback due to their adjustable duration [5], in contrast to the rest, which have fixed duration (e.g, spearcon, speech). Different audio samples were available, and participants were able to freely map specific unwanted words to the sound they preferred.

*Results.* As depicted in Figure 4 (a), participant feedback suggested a trend: shorter durations led to less distraction, while longer durations improved identifiability. Preferences were similar for the 0.5s and 1s durations, but the 2s duration seemed less preferred. Although the 1s duration seemed to offer a balance between minimizing distraction and ensuring identifiability, durations shorter than 1 second could still be effective for heightening awareness of unwanted words, provided that preference is maintained.

## 4.4 Pilot 2: Feedback Type

*Design.* We evaluated five auditory feedback types: one speech-based (i.e., the utterance of the unwanted word) and four non-speech variants (earcon, spearcon, spindex, and lyricon). Due to their inability to adequately represent specific unwanted words with suitable metaphoric mappings [10, 27], auditory icons were not considered. The earcons were set to a duration of 1s. For Lyricon—typically ranging from 0.5 to 2 seconds—were filtered to

include only variants under 1 second in duration, according to Pilot 1's results. The other feedback[4] types adhered to their original design methodologies [37, 39, 83]. For example, Table 1 shows the five auditory feedbacks for the word "like" selected by a participant. Corresponding audio samples are available as supplementary material (see META Appendix) and in the accompanying video. For the non-speech conditions, we collected the most common non-speech auditory feedback and allowed participants to map the unwanted words they wanted to reduce with the specific auditory samples based on their selection (e.g., preferences) before the actual session.

Table 1. An example of five auditory feedback types selected by a participant in Pilot 2. ★ indicates feedback types that are customizable to participants and † indicates feedback types that are common to all participants.

| Unwanted Word | Feedback Type | Example |
|---|---|---|
| Like | Earcon★ | A brief "buzz" sound intended to alert the user. |
| | Lyricon★ | A musical lyric that plays "DaDi" to provide rhythmic auditory feedback. |
| | Spindex† | A distinct sound of the letter "L" designed to be recognizable to the user. |
| | Spearcon† | An auditory feedback that involves speeding up the pronunciation of the word "like." |
| | Speech† | A playback of the word "like" itself, used as a direct auditory feedback. |

*Results.* As shown in Figure 4 (b), participant feedback identified spearcon as the most favorable overall[5]. Speech feedback ranked second in positivity but was sometimes perceived as mocking by users due to its direct delivery, leading to occasional annoyance. Although lyricon, spindex, and earcon were perceived as less annoying and less distracting, they were perceived as less intuitive, requiring users to remember specific mappings to unwanted words.

### 4.5 Pilot 3: Feedback Timing

*Design.* Based on prior findings, spearcon was selected as the feedback type. We tested six delay times: 0.5s, 1s, 2s, 3s, 4s, and 5s, focusing on specificity and preference as the primary measures. This is because spearcon supports higher intuitiveness, memorability, and privacy, among other factors (see previous pilot). The delays were implemented by adding a lag in the system whose duration was the tested delay minus 0.5 seconds, which corresponds to the minimum delay time in the Wizard of Oz setup, according to our tests.

*Results.* As shown in Figure 4(c), participant responses indicated a general trend where longer delay times led to diminished specificity and preference. Specifically, delay times above 2s significantly impacted specificity, leading to confusion among participants about the feedback's relevance to specific unwanted words, which in turn negatively affected their preference. Consequently, the results suggest that a delay time of less than 2 seconds is preferable.

## 5 WSCoach: Wearable Real-Time Auditory Feedback System Design

Informed by insights from our pilot studies, we developed *WSCoach* (**W**earable **S**peech **Coach**), a proof-of-concept wearable system designed to minimize the use of unwanted words in daily conversations through (near) real-time auditory feedback. While the implementation is relatively simple, *WSCoach* provides a means to verify whether such (near) real-time feedback is feasible and effective. *WSCoach* utilizes spearcons for feedback, chosen for their effectiveness in making unwanted words easily identifiable.

[4]Note: All non-speech feedback had durations less than 1 second, while the unwanted word determined speech feedback duration.
[5]An average score was calculated for all metrics.

*WSCoach* supports Bluetooth-enabled auditory devices (e.g., smart glasses) equipped with built-in microphones. We selected Huawei Eyewear (Version 1) for this study due to its technical compatibility and availability. The Huawei Eyewear typically provides approximately 4.5 hours of continuous talk time and up to 6 hours of continuous audio playback when operated at 60% volume.[6] Actual battery performance may vary based on usage patterns and environmental conditions. To evaluate power consumption, we conducted a pilot study with four participants ($N = 4$), assessing *WSCoach*'s battery performance during full-day natural conversations. The device, fully charged at the beginning, operated continuously in the background to detect filler words. The average battery life was 341.5 minutes (SD = 54.8 min, approximately 5.7 hours). The device uses Bluetooth 5.0, which supports a maximum data rate of 2 Mbps[7], and connects to a computer using the default sub-band codec (SBC) commonly employed in wireless headphones. SBC latency typically ranges from 150 to 250 ms[8], depending on transmission distance and environmental conditions.

*WSCoach* is implemented in Python (v3.9) and comprises three main modules: *Record*, *ASR* (automatic speech recognition), and *TTS* (text-to-speech). The source code is publicly available at https://github.com/Synteraction-Lab/WSCoach. The system runs on a desktop PC equipped with a GeForce RTX 3080 Ti GPU to enable low-latency processing.

The *Record* module captures audio using PyAudio[9] (v0.2) and interfaces with auditory device via Bluetooth. It integrates with background noise reduction tools such as NVIDIA RTX Voice[10], which employs GPU-based AI filtering to suppress background sounds during input capture. Comprehensive real-world evaluations[11] demonstrate that RTX Voice effectively reduces keyboard noise, background speech, and loud environmental sounds such as vacuum cleaners, while maintaining high voice quality.

The *ASR* module transcribes speech using the Faster-Whisper[12] (v0.7) multilingual recognition engine. Although the current setup targets detection of filler or unwanted words in English, the GUI allows users to select from other supported languages.

The *TTS* module closes the feedback loop by playing pre-generated spearcons through the smart glasses upon detection of unwanted words. Spearcons are synthesized using pyttsx4[13] (v3.0) and compressed to 40% of their original duration, following Walker et al. [83], to ensure brevity and urgency.

Overall, the system achieved an average word detection latency of 0.81 seconds (SD = 0.24, range = 0.38–1.95), followed by an average spearcon playback time of 0.28 seconds (SD = 0.08, range = 0.14–0.46), resulting in a total feedback delivery time of approximately 1.1 seconds on average.

## 6 Evaluation Study: Empirical Comparison between WSCoach and Orai

To evaluate the effectiveness of *WSCoach*, we ran a comparative study using the *Orai* mobile application (https://orai.com/) as a baseline. This commercial application offers retrospective feedback on unwanted words after a conversation. As we will explain later, it represents the status-quo approach for reducing the occurrence of unwanted words. As we are interested in the effectiveness of each system in both short-term and longer-term usage, we included two phases for our study: a training phase and a post-training phase. These two phases are detailed later in Sec 6.3.

[6] https://web.archive.org/web/20221007024433/https://consumer.huawei.com/en/wearables/huawei-eyewear/
[7] https://www.iotforall.com/bluetooth-5-iot
[8] https://www.rtings.com/headphones/tests/connectivity/bluetooth-connection
[9] https://pypi.org/project/PyAudio/
[10] https://www.nvidia.com/en-us/geforce/guides/nvidia-rtx-voice-setup-guide/
[11] https://www.techpowerup.com/review/nvidia-rtx-voice-real-life-test-performance-benchmark/3.html
[12] https://github.com/systran/faster-whisper
[13] https://pypi.org/project/pyttsx4/

The primary research questions (RQs) guiding our investigation were as follows:

- **RQ1:** Does *WSCoach* reduce the occurrence of unwanted words during the training phase, demonstrating short-term effectiveness?
- **RQ2:** Does *WSCoach* reduce the occurrence of unwanted words in the post-training phase, indicating lasting effectiveness?
- **RQ3:** Does *WSCoach* outperform *Orai* during the training phase?
- **RQ4:** Does *WSCoach* outperform *Orai* in the post-training phase?

Furthermore, we were interested in exploring the following secondary (exploratory) research questions:

- **RQ5:** How does *WSCoach* affect users' self-awareness regarding unwanted words?
- **RQ6:** How does *WSCoach* impact the overall quality of conversation?

We pre-registered our experiment design, research questions, hypotheses, and analysis plan in order to increase the reliability of our findings [60] (see META Appendix).

## 6.1 Participants

We recruited 24 participants (P1-P24, 12 male, 12 female, mean age = 23, SD = 3.5) from our university, none of whom took part in our pilot studies. All participants reported in a recruitment questionnaire that they have no auditory or visual impairment, that they want to reduce unwanted words in their daily conversations, and that they have professional-working fluency in English. They were compensated 7.31 USD per hour for their time.

The sample size of $N$=24 was chosen ahead of time, considering that it would give us a power of 0.7 to detect a standardized effect size of $d$=1 (conventionally referred to as a large effect size) using an independent-sample t-test, as calculated with G*Power. Although such a sample size is insufficient to detect small differences in effectiveness between our two experimental conditions, we were constrained by the time and financial costs of running participants.

## 6.2 Apparatus

The study required participants to converse with the experimenter using either the *WSCoach* (Sec 5) or the *Orai* system. For the *Orai* system, conversations were facilitated using the *Orai* mobile app on a Redmi K40 smartphone. All conversations were audio-recorded with a Huawei P10 to allow for post-analysis.

*Orai* was selected as the baseline of comparison due to its reputation as an effective AI speech coach, offering comprehensive training in speaking skills. It is a highly popular app, evidenced by over 100,000 downloads and a 4.1-star rating on Google Play. The app features post-conversation analysis, including audio transcriptions highlighting unwanted words, their frequency, and playback capability for in-depth self-evaluation (See Appendix A.5).

## 6.3 Study Design

The study followed a between-subjects design, as shown in Fig 5. Participants were randomly assigned to one of two groups: *WSCoach* or *Orai*. This assignment was based on a pre-generated, shuffled list that determined the condition of each participant according to the order in which they completed the recruitment questionnaire.

All conversation sessions were conducted in a controlled laboratory setting, ensuring minimal environmental interference such as noise or other distractions. All participants adhered to their scheduled times and completed the entire experiment.

Each participant engaged in eight conversation sessions across 15 days, for a total of about four hours of participation. The sequence of sessions is summarized in Figure 5 for each experimental condition. We provide a brief overview here, with more details in the following subsection. The first conversation took place without any technology assistance. Then, the training phase consisted of six conversation sessions with the system (*WSCoach*

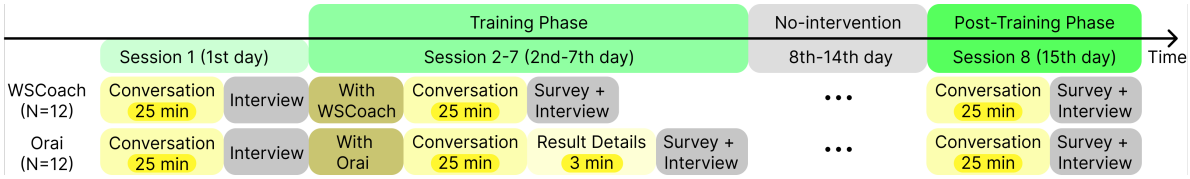


Fig. 5. Experiment design used in the comparative evaluation, showing the sequence of experimental sessions for the *WSCoach* experimental group (top row) and the *Orai* group (bottom row).

or *Orai*), on consecutive days. This number was chosen based on prior studies that apply behavioral interventions across six sessions [24, 30]. Each conversation lasted about 25 minutes, measured by a timer starting at the beginning and stopping once the time had elapsed, recording the total duration. This choice was inspired by a well-known time management method that recommends working in 25-minute segments [14]. The training phase was followed by a no-intervention period during which participants did not use any system. Finally, participants attended a final session to assess their improvement.

## 6.4 Task and Procedure

We now go through the phases shown in Figure 5 and detail the procedure.

**First session**. After giving consent and receiving the briefing, the participant was seated at a table (see Figure 1) and had a first conversation with the experimenter (one co-author, always the same), without any technological assistance. The goal of this first session was to establish a baseline count of unwanted words, and to help participants familiarize themselves with the settings. The conversation was guided by the experimenter, based on topics derived from IELTS Speaking Questions (https://ielts.org/), which cover a range of familiar subjects. Each session used a different set of topics (see META Appendix). The experimenter assumed a facilitator role by listening attentively, aligning with participants' viewpoints, and steering the conversation when it began to stall. They were also trained to avoid speaking unwanted words and to suspend system detection when the participant was not speaking. After the conversation, the participant was interviewed to identify which unwanted words they wished to reduce.

**Training phase**. This phase took place from day 2 to day 7. On the second day, the participant was introduced to the system that was randomly assigned to them (*WSCoach* or *Orai*). After familiarization with the system, the participants engaged in a conversation session, assisted by the system. They kept using the same system for the next five conversation sessions (days 3 to 7). In the *WSCoach* group, participants put on smart audio glasses with their lenses removed before each conversation. Those already wearing spectacles were invited to remove them or place smart audio glasses over them. In the *Orai* group, participants were handed a smartphone with *Orai* installed, and were invited to open the app and place the phone on the table to monitor the conversation. They then had to spend at least 3 minutes reviewing their use of unwanted words with the *Orai* application. Additionally, in both groups, participants completed a questionnaire and participated in an interview about their experiences and the system's pros and cons (see Sec 6.6 on Measures for more details). In the last training session (the 7th session), participants completed an additional questionnaire and participated in a 15-minute interview about their perceived improvement during the training phase. The complete training phase interview questions are available in META Appendix.

**No-intervention period**. The training phase was followed by an 8-day break during which participants were not involved in any sessions and did not use any system.

**Post-training phase** On the 15th day, participants engaged in their final conversation with the experimenter. Afterward, they completed a post-training phase questionnaire and were interviewed about their perceived improvement. The complete post-training phase interview questions are available in META Appendix

## 6.5 Post-Experiment Data Processing and Analysis

*Quantitative Coding.* Post-experiment, all conversation recordings and post-interview recordings were anonymized and transcribed with the help of Deepgram (https://deepgram.com/). We reviewed all transcripts for accuracy and redacted any content that might disclose the experimental condition (*WSCoach* or *Orai*), for the purpose of the subsequent blind coding. Two independent coders, who did not attend the interviews and were naive about the purpose of the research, were provided with the anonymized conversation transcripts and a list of unwanted words for each conversation (designated by the participant in the first session). The coders were tasked with counting the occurrences of unwanted words in each conversation and measuring speaking time, i.e., the amount of time the participant speaks. We assessed inter-coder agreement for the counts of unwanted words and the amount of time participant speaks using Cohen's Kappa [15], obtaining $\kappa = 0.85$ and $\kappa = 0.81$ separately, both reflecting high agreements. Disagreements were resolved through discussion between the two coders.

*Qualitative Coding.* Moreover, we used thematic analysis [7] to analyze the qualitative data (e.g., interview). First, we (two co-authors) reviewed the qualitative data records multiple times. Then, we identified the participants' responses, grouped the answers, and incorporated them into themes and sub-themes. We finalized our themes and discussed the different speech intervention systems in detail, along with guidelines for real-time auditory intervention. We report the qualitative results later in the paper.

## 6.6 Measurements

The effectiveness of reducing unwanted words and the quality of conversation were evaluated using objective and subjective measurements, as summarized in Table 2. In addition, participants' perceptions of real-time auditory feedback and delayed memory were captured in the interviews.

To help address statistical multiplicity [3, 45], we separate our measurements into *primary measurements*, which will serve to answer our main research questions, and *secondary measurements*, which are more exploratory. Additionally, we introduced a set of *supplementary measurements*, which were not pre-registered and added after peer review, to address reviewers' questions and comments.

*6.6.1 Primary Measurements.* We have two primary measurements, both of which are defined based on a metric we call the *normalized frequency of unwanted words*. We refer to the normalized frequency of unwanted words in session $n$ as $S_n = \frac{F_n}{F_1}$, with $1 \leq n \leq 8$, and $F_n$ being the *frequency of unwanted words* in session $n$, defined as the ratio between the number of unwanted words in session $n$ and the participant's total speaking time in that session. Accordingly, $S_1 = 1$ for all participants. Our first primary measurement is the rate of change in $Sn$ from the 1st to the 7th session. The way the seven measurements of normalized frequency (one per session) are aggregated into a single rate measurement will be explained in more detail in the analysis section. This measurement captures the participant's improvement during the training phase and will serve to answer research questions RQ1 and RQ3 (see Sec 6). Our second primary measurement is $F_8$, which captures the long-term improvement as measured in the post-training phase, and it will serve to answer RQ2 and RQ4.

*6.6.2 Secondary Measurements.* For the training phase, we evaluated several aspects to determine *WSCoach*'s impact on conversation quality and user experience. We recorded *Self-rated Conversation Quality* to see if the system made conversations feel unnatural. We measured the *Distraction of Feedback* to assess if participants were distracted by the system. Frequent auditory feedback might negatively affect participants' confidence,

Table 2. The measurements used in Study 2.

| | | |
|---|---|---|
| **Independent Variables** | **Platform** | • Glasses<br>Using *WSCoach* for reducing unwanted words<br>• Smartphone<br>Using *Orai* for reducing unwanted words |
| **Measurements (Dependent Variables)** | **Primary Measurements** | • *Improvement during the training phase,* measured as the rate of change in the normalized frequency of unwanted words $S_n$, from the 1st to the 7th session (see Sec 6.6).<br>• *Improvement at the post-training phase*, defined as the normalized frequency of unwanted words at session 8 (Sec 6.6). |
| | **Secondary Measurements** | • *Training phase – Confidence in Speaking* (1-7): "This system made me feel more confident in my conversations."<br>• *Training phase – Awareness of Speaking Unwanted Words* (1-7): "This system helped me become more aware of my unwanted words during conversations."<br>• *Training phase – Self-rated Conversation Quality* (1-7): "My quality of conversation has improved with the help of this system."<br>• *Training phase – Distraction of Feedback* (1-7): "The system distracted me from having a natural conversation."<br>• *Training phase – Raw NASA TLX* (0-100) for Being Aware of Unwanted Words<br>• *Post-training phase – Confidence in Conversation* (1-7): "I felt confident in maintaining the reduction of unwanted words without the aid of the system."<br>• *Post-training phase –Awareness of Avoiding Unwanted Words* (1-7) = "I found myself consciously avoiding unwanted words." |
| **Supplementary Measurements** | | • *Normalized Ratio of Unwanted Words*: The total number of unwanted words divided by the total number of words participant spoke. |

so we measured *Confidence in Speaking*." Awareness of unwanted words indicates participants' improvement, so we measured *Awareness of Speaking Unwanted Words*. All these metrics were measured on a 7-point Likert scale. Additionally, we evaluated the cognitive load using the perceived task load (NASA TLX) [29] to assess the workload involved. For the post-training phase, we collected *Confidence in Conversation* to evaluate the system's lasting effect on users' confidence and *Awareness of Avoiding Unwanted Words* to measure the lasting self-awareness regarding unwanted words. These measurements in the training and post-training phases will serve to answer RQ5 and RQ6.

*6.6.3 Supplementary Measurements.* We calculated the *Normalized Ratio of Unwanted Words* as an alternative metric to analyze the improvement during the training phase and the post-training phase. For each session $n$, it is defined as $S'_n = \frac{R_n}{R_1}$, with $1 \leq n \leq 8$, and $R_n$ being the ratio of unwanted words in session $n$, defined as the ratio between the number of unwanted words in session $n$ and the number of participant's total words in that session. This new metric has the advantage of not being affected by changes in speech rate.

## 6.7 Statistical Analysis

*6.7.1 Planned Analysis.* The analyses reported here were planned and preregistered before data collection (see META Appendix).

***RQ1:*** *Does WSCoach reduce the occurrence of unwanted words during the training phase?* For each participant in the *WSCoach* group, we performed a linear regression with normalized frequency $S_n$ as the dependent variable and session number $n \in [1..7]$ as the independent variable. We used the regression slope to measure the rate of change in unwanted words across the training phase for each participant. We then computed the mean slope for

the entire *WSCoach* group and performed a one-sample $t$-test with zero (horizontal slope, meaning no change overall) as the null hypothesis. From this $t$-test we derived a $t$-based 95% confidence interval for the mean slope and a $p$-value, capturing the amount of evidence in favor of an overall reduction of unwanted words during the training phase.

***RQ2:*** *Does WSCoach reduce the occurrence of unwanted words in the post-training phase?* We performed a one-sample $t$-test on the mean normalized frequency $S_8$ (normalized frequency at the 8th session, i.e., after about 15 days) for the *WSCoach* group, with 1 as the null hypothesis (meaning no change). This $t$-test allowed us to assess whether there was a reliable overall improvement after the no-intervention period compared to the very first session.

***RQ3:*** *Does WSCoach outperform Orai during the training phase?* For each participant in the *Orai* group, we performed the same linear regression as we did for the *WSCoach* group (see RQ1). We then performed an independent-sample $t$-test to see if there is a reliable difference between the mean slopes of the two groups.

***RQ4:*** *Does WSCoach outperform Orai in the post-training phase?* We performed an independent-sample $t$-test to compare the mean $S_8$ between the *WSCoach* group and the *Orai* group.

*6.7.2 Secondary Analyses of Primary Measurements.* In this subsection and all following subsections, all analyses are post-hoc and not pre-registered. Therefore, any finding should be taken as tentative.

*Normalized Frequency S7.* Our regression method is one way of operationalizing improvement in the learning phase (RQ1), but another way is to compare the last training session with the initial session, i.e., by looking at the normalized frequency $S_7$. We performed an independent-sample $t$-test to compare the mean $S_7$ between the two groups.

*Robustness Check Using S'.* In addition, we introduced a new metric for measuring the occurrence of unwanted words: $S'_n$, the *Normalized Ratio of Unwanted Words*, defined in Sec 6.6. As a robustness check, we re-analyzed our data using $S'_n$ instead of $S_n$ and re-assessed our primary research questions (RQ1 to RQ4).

*6.7.3 Analyses of Secondary Measurements.* For all Likert items, we computed the mean response (range 1–7) and its $t$-based 95% CI for each group. We also computed the $p$-value for the Mann-Whitney $U$ test of the difference between groups.

All statistical analyses were conducted using SPSS (Statistical Package for the Social Sciences).

## 7 Evaluation Study: Results

Overall, both *WSCoach* and *Orai* helped reduce participant-identified unwanted words. While *WSCoach*'s real-time feedback caused more distraction than *Orai*'s retrospective approach, it led to greater reductions in unwanted words and higher user confidence. We report our findings below, distinguishing between pre-registered and exploratory analyses in line with methodological guidelines [45].

### 7.1 Results of the Planned Analysis.

We report $p$-values but interpret them as continuous measures of evidence, using .05 as a rough landmark instead of a cut-off [4].

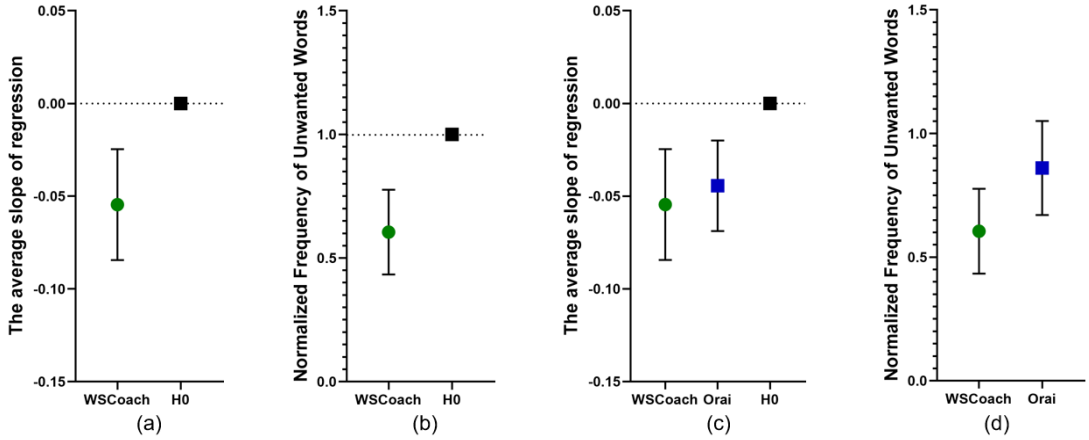


Fig. 6. The planned analyses of primary measurements for the evaluation study. Lower is better. Error bars are 95% confidence intervals. Black squares and dashed lines show the null hypothesis.

***RQ1:*** *Does WSCoach reduce the occurrence of unwanted words during the training phase?* As shown in Figure 6 (a), the mean slope was $M = -0.05$[14], 95% CI [-0.08, -0.03], $t(11) = 4.02$, $p = .002$. Therefore, we have strong evidence that the *WSCoach* led to an overall improvement during the training phase.

***RQ2:*** *Does WSCoach reduce the occurrence of unwanted words in the post-training phase?* As shown in Figure 6 (b), the mean normalized frequency was $M = 0.60$, 95% CI [-0.57, -0.22], $t(11) = 5.07$, $p = .0004$. Therefore, we have strong evidence that *WSCoach* was effective and lasted up to the post-training phase.

***RQ3:*** *Does WSCoach outperform Orai during the training phase?* As shown in Figure 6 (c), the difference in mean slopes (*Orai*−*WSCoach*) was 0.01, 95% CI [-0.03, 0.05], $t(22) = 0.58$, $p = .57$. Therefore, we have no evidence that *WSCoach* outperformed *Orai* during the training phase. Additionally, the mean slope and its 95% confidence interval for each of the two groups show that the effectiveness estimates are comparable (i.e., *Orai* is effective as well).

***RQ4:*** *Does WSCoach outperform Orai in the post-training phase?* As shown in Figure 6 (d), the difference between the two means (*Orai*−*WSCoach*) was 0.26, 95% CI [0.01, 0.50], $t(22) = 2.2$, $p = .04$. Therefore, we have evidence that *WSCoach* outperformed *Orai* in the post-training phase. Additionally, the mean normalized frequencies and their 95% confidence intervals suggest that *WSCoach* is likely more effective in the long run (the lower the value, the better).

## 7.2 Results of the Secondary Analyses of Primary Measurements

*Normalized Frequency S7.* As shown in Figure 7 (a), the difference between the two means (*Orai*−*WSCoach*) was 0.24, 95% CI [0.02, 0.47], $t(22) = 2.2$, $p = .04$. Therefore, this way of analyzing improvement during the learning phase might be more sensitive, but a replication is needed to confirm it.

*Robustness Check Using S'.* The conclusions are unchanged. Results are reported in Figure 7 (b)–(e), and can be compared to our original results in Figure 6 (a)–(d):

[14]For all our reported slope values, the dependent variable is $S_n$ (unitless), and the independent variable is the session. Therefore, the slopes represent the change in the unwanted word ratio per session, but there are no units following them.

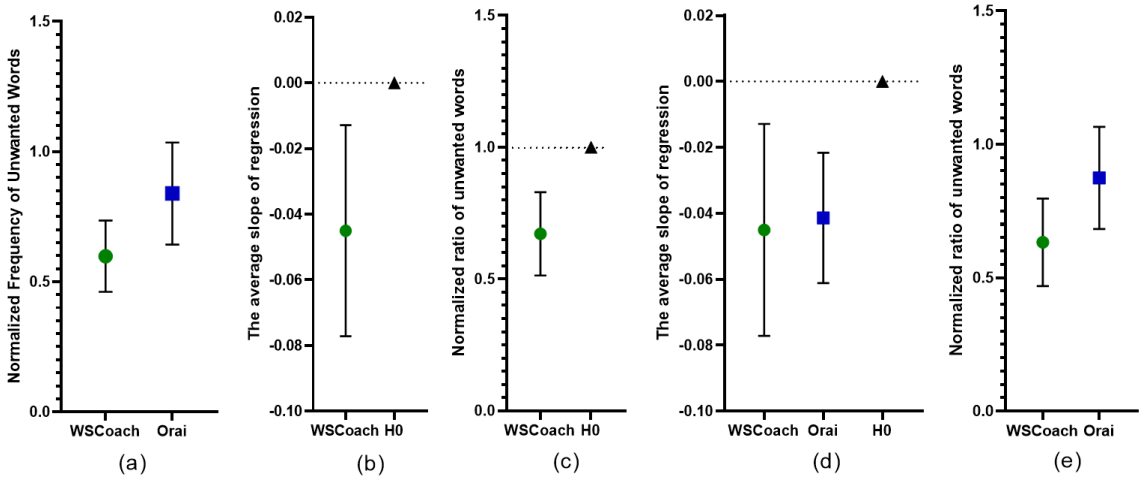


Fig. 7. The secondary analyses of primary measurements for the evaluation study. Lower is better. Error bars are 95% confidence intervals. Black triangles and dashed lines show the null hypothesis.

*RQ1:* As shown in Figure 7 (b), the mean slope was $M = -0.05$, 95% CI [-0.08, -0.01], $t(11) = 3.09$, $p = .01$. Therefore, we still have strong evidence that the *WSCoach* led to an overall improvement during the training phase.

*RQ2:* As shown in Figure 7 (c), the mean normalized ratio was $M = 0.67\%$, 95% CI [-0.49, -0.17], $t(11) = 4.57$, $p = .0008$. Thus we retain the strong evidence that *WSCoach* was effective up to the post-training phase.

*RQ3:* As shown in Figure 7 (d), the difference in mean slopes (*Orai-WSCoach*) was 0.004, 95% CI [-0.03, 0.04], $t(11) = 0.21$, $p = .83$. Therefore, like before, we have no evidence that *WSCoach* outperformed *Orai* during the training phase. Additionally, the mean slope and its 95% CI for each of the two groups show that the effectiveness estimates are comparable (i.e., both methods are effective).

*RQ4:* As shown in Figure 7 (e), the difference between the two means (*Orai-WSCoach*) was 0.24, 95% CI [0.003, 0.48], $t(22) = 2.1$, $p = .05$. Therefore, we still have evidence that *WSCoach* outperformed *Orai* in the post-training phase.

### 7.3 Results of the Secondary Measurements.

Results are reported in Figure 8 (a) to (g). Here, we report the full results for the two questions for which we have evidence of a difference—the full analysis is available in META Appendix.

*Distraction of Feedback.* As shown in Figure 8 (d), the mean score was 4.1, 95% CI [3.0, 5.2] for *WSCoach* and 1.8, 95% CI [1.1, 2.4] for *Orai*, $U = 117$, $p = .003$. Therefore, we have good evidence that *WSCoach* generated more distraction on average than *Orai* during conversations.

*Confidence in Conversation (post-training phase).* As shown in Figure 8 (f), the mean score was 5.3, 95% CI [4.9, 5.6] for *WSCoach* and 4.3, 95% CI [3.8, 4.9] for *Orai*, $U = 28$, $p = .01$. Therefore, we have evidence that *WSCoach* made people feel more confident in their conversation than *Orai* after the training was over.

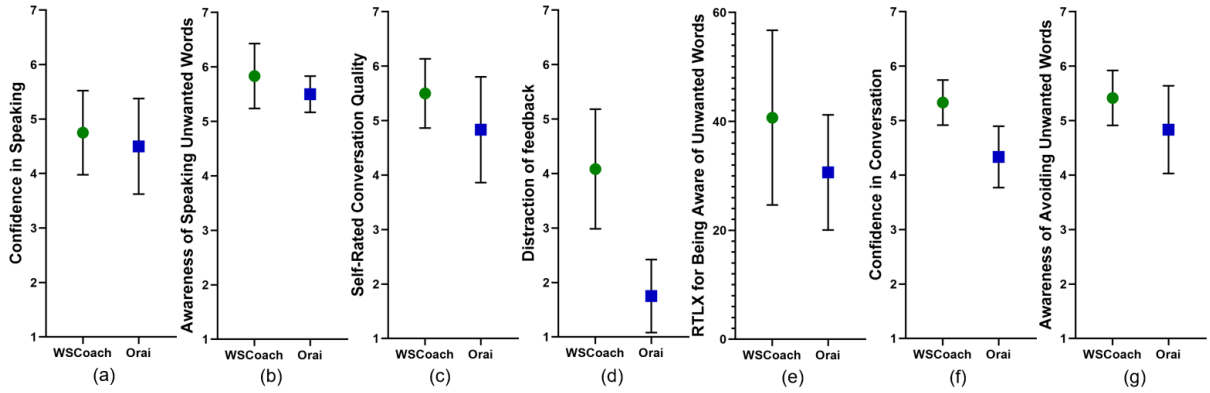


Fig. 8. The analyses of secondary measurements for the evaluations study. Higher is better, except for (d) and (e), where lower is better. Error bars show 95% confidence intervals.

## 7.4 User Feedback and Discussion

Based on the interview recording analysis, we discuss the user experience on two systems.

*Established a Self-Monitoring Mechanism.* Participants' feedback suggests that *WSCoach* helped establish an active self-monitoring mechanism during the training phase which helped users in reducing unwanted words: —"*Whenever I hear the auditory feedback, I consciously allow myself to slow down and reflect on my words. I integrate the auditory feedback and the instruction that advises me to slow down unconsciously when approaching the expression of unwanted words during conversations.* (P7)" Furthermore, our study shows that 25-minute training sessions over a 6-day period were sufficient to have a 40% reduction in unwanted words on average in the post-training phase.

*Improved Awareness and Lasting Impact.* Interestingly, after 3-4 sessions of using *WSCoach*, participants noted improved awareness during their daily communication even when not using *WSCoach*, highlighting the potential of *WSCoach* to yield lasting improvements. For instance, Participant P13 mentioned on their 4th day that, "*During daily conversations, I have started noticing instances where I use unwanted words, prompting me to pay extra attention to monitor these words... when I'm about to utter these words, I feel a distinct 'real-time auditory feedback' echoing those words in my mind... it [WSCoach] has helped develop a pattern to avoid speaking those unwanted words.*" Such behavioral changes are a result of operant conditioning [76], which posits that behaviors followed by negative consequences are less likely to be repeated. In our case, the real-time auditory interventions of *WSCoach* acted as negative feedback – an unfavorable consequence contingent on the use of unwanted words. When users speak more unwanted words, they receive increased auditory feedback, creating a negative reinforcement loop. P2's preference for reduced auditory feedback underscores a common inclination to minimize negative consequences by reducing the utterance of unwanted words, —"*When I initiate a conversation, I consciously remind myself to aim for fewer auditory feedback instances. I prefer the system to detect fewer unwanted words in my speech. Experiencing more auditory feedback makes me feel uneasy, as it suggests a lack of improvement in my communication habits.* (P2)"

As highlighted earlier, our results indicate that *WSCoach* outperforms retrospective feedback (i.e., *Orai*) after the training is over. The higher benefits of *WSCoach* can be primarily attributed to *WSCoach* pinpointing instances of unwanted words in real-time and enhancing the awareness of unwanted words during conversations. P8 expressed, "*WSCoach tells me what unwanted words I spoke, which helps me identify the locations of unwanted*

*words in my speech during conversations. This enables me to proactively minimize them, structuring my sentences more deliberately before speaking.*" In contrast, the retrospective feedback offered by *Orai* post-conversation was found to be less actionable, resulting in a diminished immediate impact and a weaker context connection for users regarding the identified unwanted words. In addition to the real-time awareness, *WSCoach* affords opportunities to users to proactively practice reducing the occurrence of unwanted words, in contrast to retrospective feedback, —"*I am aware of the number of unwanted words I used after the conversation [using Orai], but I believe it's not enough. It doesn't guide me on how to reduce the unwanted words while conversing.* (P1)"

*Impact on Distraction and Cognitive Load.* One important aspect of *WSCoach* is its variable impact on perceived distraction during conversations, particularly in the early stages of use. While most participants (8/12) reported acclimating to the auditory feedback by the third or fourth day—eventually no longer finding it disruptive—this was not universal. Notably, one participant (P19) consistently found the system distracting, which paradoxically led to an increase in their use of unwanted words. As they described: —"*I believe the reason this intervention might not be friendly to me is that my thoughts tend to flow continuously, and I can be quite sensitive to negative feedback. So, when I'm interrupted, I need to reorganize my thoughts; in such situations, my use of unwanted words might increase.* (P19)" This experience suggests that while real-time feedback can raise awareness and encourage behavior change, it may also disrupt cognitive flow—particularly for users who are sensitive to interruption or who rely on continuous verbal expression. For such individuals, real-time feedback may inadvertently increase cognitive load and hinder performance, at least initially.

Cognitive load is a critical factor in behavior change interventions, especially when targeting deeply ingrained habits [42]. *WSCoach* introduces some initial effort: users must attend to both their speech and the system's feedback, which can momentarily disrupt the natural rhythm of conversation. Nonetheless, most participants found that this disruption subsided after 2–3 days, as the feedback became less intrusive and more intuitive. This adaptation curve aligns with habit formation research, which emphasizes the role of repetition and contextual cues in shifting behaviors from effortful to automatic. As Lally et al. [43] note, such transitions often unfold over weeks or months, highlighting the need for sustained support. In our study, both *WSCoach* and the control condition (*Orai*) led to short-term reductions in unwanted word use. However, these gains tended to diminish by Day 15 when the intervention was withdrawn, suggesting that while immediate benefits are achievable, long-term change may require ongoing or periodic reinforcement.

Ultimately, we view the initial cognitive load not as a deterrent, but as a natural part of the behavior change process. *WSCoach* is designed to minimize disruption through brief, subtle spearcons, and our findings suggest it is generally well tolerated. Still, long-term usability and cognitive effort remain important considerations.

This diversity in user responses highlights the need to account for individual differences in the design and application of real-time auditory interventions. A viable solution could involve adaptive features allowing users to tailor the feedback's intensity or frequency to their preferences, thus harmonizing the intervention with their cognitive processes. Additionally, a gradual feedback intensification strategy over initial sessions may facilitate a smoother adaptation, preventing users from feeling overwhelmed and promoting gradual adjustment. Personalizing interventions is key to accommodating diverse user needs, thereby maximizing efficacy and minimizing potential distractions.

## 8 General Discussion

The *WSCoach* prototype demonstrates how immediate auditory intervention during conversational moments can effectively modify habitual speech behaviors in everyday interactions. Building such real-time intervention systems requires careful consideration of feedback types and real-time auditory intervention design. Below, we outline key aspects that future systems should consider and refine for effective digital speech behavior interventions.

### 8.1 Real-Time vs. Retrospective Feedback for Speech Intervention

Our research highlights the impact of real-time auditory feedback on effective speech-behavioral intervention. By delivering immediate auditory feedback, users become more aware of specific speech patterns, enabling them to recognize and address unwanted behaviors in context. This immediacy allows for quick reflection and self-correction, creating a reinforcement loop that promotes sustainable speech behavior modification. During our experiments, participants responded to auditory feedback by considering recently used unwanted expressions and adjusting their speech accordingly. This approach contrasts with retrospective feedback, such as that offered by public speaking tools like *Orai*, where users can review and analyze their performance post-speech in a focused environment. While useful for skill-building in structured, single-focus tasks, retrospective feedback lacks the immediacy essential for spontaneous speech correction within interactive dialogue. In dynamic, multitasking environments, real-time feedback is uniquely suited for enabling users to identify and regulate behaviors such as filler word overuse without interrupting conversational flow. This approach aligns well with the demands of natural conversations, enhancing real-time communication skills by encouraging self-awareness and facilitating rapid improvements.

### 8.2 Privacy Considerations for Real-Time Speech Intervention Systems

Privacy is a central concern in the design and deployment of real-time speech feedback systems like *WSCoach*. Although the system processes only **audio-only** input (i.e., no visual data), its always-on nature and operation in social environments raise important privacy considerations [22] for both users and bystanders.

*User Privacy.* *WSCoach* is designed to perform speech detection in real time and discard raw audio immediately after processing. Only minimal metadata (e.g., frequency of unwanted word usage) is retained. Nevertheless, several participants (3/12) expressed concerns about the possibility of sensitive speech being unintentionally processed, even when the **audio itself is not stored**. While *WSCoach* avoids long-term recording, its continuous operation still led some users to perceive it as a form of surveillance. However, real-time auditory feedback was generally regarded more favorably than retrospective, phone-based recordings. More participants (6/12) viewed the latter as more invasive due to the potential for long-term storage and external access.

As edge computing capabilities continue to advance, future iterations may shift more of the processing directly onto the glasses themselves, further reducing the need for data transmission and minimizing the risk of data exposure [52, 89]. Additionally, user-facing privacy controls—such as those modeled on Apple's App Tracking Transparency framework [34]—can enable fine-grained control over when processing occurs and what data is retained.

Moreover, several participants (4/12) expressed a preference for subtle auditory feedback to avoid drawing attention from bystanders and to maintain social appropriateness. These insights underscore the importance of delivering feedback discreetly and in a privacy-preserving manner—ideally through speaker-only audio using bone conduction or directional sound.

*Bystander Privacy.* While users may consent to speech processing, bystanders may not. This introduces a distinct ethical challenge. One participant (P14) specifically noted discomfort using the system around friends or family, expressing concern that it could inadvertently process sensitive portions of others' conversations. Unlike some mobile apps that store entire conversations, *WSCoach* neither stores audio nor attempts to associate it with specific individuals. Nevertheless, the mere presence of a listening device can raise concerns about passive surveillance.

To address this, future versions of *WSCoach* will explore integrating speaker diarization and voice activity detection techniques to isolate the wearer's voice. Combined with spatial filtering through multi-microphone

arrays in smart glasses, these techniques—based on recent advances in wearable voice segmentation and beamforming—could help ensure that feedback is triggered only by the user's speech. This capability is crucial for maintaining bystander trust and enabling ethical, socially acceptable deployment.

By limiting data collection, enabling on-device processing, and supporting selective voice detection, *WSCoach* aims to balance utility with strong privacy safeguards. These considerations are essential not only for user adoption but also for the responsible integration of speech feedback technologies into everyday social contexts.

## 8.3 What Kind of Real-Time Feedback Should be Provided?

Participants using *WSCoach* highlighted its effectiveness in helping them recall and reduce the use of unwanted words during conversations. Prior studies have shown that detailed feedback enhances knowledge retention and the application of knowledge [85]. The design of real-time auditory feedback is pivotal in achieving this, demanding a careful balance between the richness of information, duration, and minimization of distraction during conversations. Additionally, balancing frequent practice with user endurance is essential when designing real-time auditory feedback systems that foster lasting behavior change.

*8.3.1 Trade-off: Detailed Information and Duration.* The evaluation study (Sec 7) has shown that auditory feedback with sufficient information effectively reduces various unwanted words during live conversations. However, while providing detailed information, the duration of auditory feedback increases, which may disturb users. Our pilot study results suggest that feedback should be detailed enough to convey information about specific unwanted words spoken, ensuring it remains meaningful. Nonetheless, it should also be limited to a maximum duration of 1 second to allow users to quickly comprehend it without disrupting the flow of conversation. This balance can be achieved through the careful selection and prioritization of spearcon content for different unwanted words. Managing this trade-off is essential for optimizing the auditory feedback system's effectiveness and user experience.

*8.3.2 Trade-off: Detailed Information and Distraction.* We observed a slight distraction while using *WSCoach* due to the reception of detailed information with real-time auditory feedback during fluent conversations. Our pilot studies found that other non-speech-based feedback is less distracting than spearcons; however, spearcons provide more informative feedback. Notably, the majority of participants (11/12) valued the detailed feedback. Although it initially caused some distraction, most adapted within three sessions, after which the distraction became negligible. Given the typical conversational rate of 110-150 words per minute [49], even brief, 1-second feedback spans only about 2 to 3 words—well within the human short-term memory limit of approximately 3 to 5 chunks [17]. This brief auditory feedback minimally disrupts conversational flow, allowing users to retain context and continue interacting naturally. Therefore, while the detailed information may induce initial distraction, users generally adjust and become accustomed to it if it is well designed.

The desire to improve speaking skills stands out as a key motivator, encouraging users to adapt to potential distractions. The stronger the user's goal to achieve meaningful outcomes, the higher their capacity to endure and benefit from the inherent distractions within the habit reversal process. P19 echoed this sentiment, expressing a keen interest in improving their speaking skills and reducing awkward and unwanted words. Consequently, they do not perceive the detailed information with auditory feedback as a distraction; instead, they appreciate it because it helps them improve quickly. Understanding these factors provides a foundation for tailoring methodologies to individual preferences, fostering a more personalized and effective habit-reversal experience.

*8.3.3 Trade-off: Ubiquitous Practice and User Endurance.* As real-time auditory intervention systems become more integrated into daily life, maintaining user engagement without inducing fatigue is crucial. Ubiquitous intervention, while potentially effective, can risk user overload if not thoughtfully implemented. To mitigate

this, adaptive scheduling can allow users to control the frequency and intensity of coaching sessions or specify contexts in which the coach should activate.

## 8.4 Expanding Wearable Real-Time Auditory Interventions

*8.4.1 Incorporating Contextual Awareness.* Dynamically adjusting feedback based on user context—such as emotional state [33] or gesture cues—may enhance both user experience and system effectiveness. In the case of delivering auditory feedback for unwanted words, it is important that the system consider not only what was said, but also how and when it was said. For example, signs of frustration or stress detected through vocal tone may indicate that the user is already cognitively burdened. In such cases, reducing the frequency or intensity of feedback could help prevent additional discomfort. Similarly, nonverbal signals such as posture, facial expression, or hesitation gestures may suggest agitation or uncertainty, signaling the system to delay or soften feedback delivery.

Moreover, not all unwanted words are inherently negative. Depending on the user's context, filler words or informal expressions may serve communicative functions—such as emphasizing a point or maintaining conversational flow. Automatically flagging these without considering intent may disrupt the user unnecessarily and reduce the system's utility. Understanding the contextual role of such expressions is therefore critical for effective intervention. This adaptability is particularly important in multilingual or multicultural settings, where rigid feedback may be misinterpreted. Context-aware strategies not only enhance relevance but also align with best practices in behavior change, which emphasize tailoring interventions to individual needs. Incorporating emotional and behavioral cues into feedback design offers a promising path toward more responsive and empathetic wearable assistants.

*8.4.2 Enhancing Multilingual and Holistic Behavioral Interventions.* This study focuses primarily on English, but the core functionality of the *WSCoach* system can be extended to other languages. It detects unwanted words and delivers auditory feedback—a mechanism adaptable to any language. While the patterns of unwanted word usage may vary across languages, the underlying approach remains consistent. The system can be configured to identify language-specific unwanted expressions, ensuring that feedback effectively raises user awareness and supports behavior modification across diverse linguistic contexts.

Beyond unwanted speech, the system's core functionality can be applied to other behavioral habits. For example, it could support improvements in social behavior—such as prompting users to greet others—or help manage unconscious negative facial expressions by providing real-time auditory feedback that increases self-awareness. Making the system language-agnostic and expanding its scope to broader behavioral interventions would significantly enhance its applicability, establishing *WSCoach* as a more versatile tool for supporting holistic personal development.

# 9 Limitations and Future Work

The present research calls for more work to address its limitations, both in terms of the study and the prototype.

*System Limitations.* False detections occasionally occurred when bystanders (i.e., experimenters) used words from the user's list of unwanted terms. In such cases, the system could not reliably differentiate the user's voice from others, resulting in unintended feedback. This limitation stems from hardware constraints in most commercial smart glasses, including those used in our study, which lack spatial audio filtering or beamforming capabilities for speaker diarization. Consequently, the system could not directionally isolate the user's speech. To address this, future iterations could incorporate multi-microphone spatial filtering to better isolate the user's voice and reduce false positives.

*Implementation Limitations.* The current implementation of *WSCoach* runs on a GPU-enabled laptop connected to the smart glasses via Bluetooth, enabling high-accuracy, low-latency speech detection. While this setup supports reliable performance in typical indoor scenarios (e.g., indoor meetings or presentations with slides), it may limit portability and usability in mobile or outdoor contexts. Advances in on-device processing—such as edge AI accelerators in smartphones (e.g., Apple Neural Engine, Qualcomm Hexagon DSP)—could enable real-time inference directly on mobile or wearable devices. These developments would make it feasible to deploy *WSCoach* in more mobile or resource-constrained environments without compromising performance.

*Evaluation Limitations.* Our findings, though promising in a laboratory environment that mimicked typical indoor conditions, require further validation in longer-term and more diverse real-world contexts to enhance ecological validity. Although we minimized strict controls—allowing ambient variability akin to everyday settings—the generalizability of our results may still be constrained by limited environmental and demographic diversity. Expanding participant demographics and incorporating multilingual support are important next steps toward broadening the applicability of *WSCoach*. We also acknowledge that adapting the system to handle auditory interference in more challenging environments, such as noisy outdoor settings, will be necessary for reliable deployment across varied contexts.

## 10 Conclusion

Our study sheds light on the potential of auditory feedback as a tool for enhancing self-awareness and reducing the use of unwanted words in daily conversations. The findings underscore the importance of selecting the appropriate type, duration, and delay time of feedback to achieve targeted speech corrections. These insights offer guidance to the development of speech intervention tools and the understanding of speech behavior correction. Looking forward, future work could focus on developing a general wearable real-time behavior coach and enhance people's awareness about their improper behavior in various application scenarios.

## Acknowledgments

This research is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (Award Number: AISG2-RP-2020-016). The Guangxi Science and Technology Base and Talent Special Project (No. guikeAD23026230), CityU Start-up Grant (No. 9610677), and National Natural Science Foundation of China (No. 62462003) also provide partial support. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not reflect the views of the National Research Foundation, Singapore. We extend our gratitude to all members of the Synteraction Lab for their help in completing this project. We also thank the reviewers for their valuable feedback.

## References

[1] Fouad Alallah, Ali Neshati, Yumiko Sakamoto, Khalad Hasan, Edward Lank, Andrea Bunt, and Pourang Irani. 2018. Performer vs. observer: whose comfort level should we consider when examining the social acceptability of input modalities for head-worn display?. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3281505.3281541

[2] Drew Ashby-King, Raphael Mazzone, and Lindsey Anderson. 2021. Defining Feedback: Understanding Students' Perceptions of Feedback in the Introductory Communication Course. *Journal of Communication Pedagogy* 4, 1 (Sept. 2021). doi:10.31446/JCP.2021.1.06

[3] Ralf Bender and Stefan Lange. 2001. Adjusting for multiple testing—when and how? *Journal of clinical epidemiology* 54, 4 (2001), 343–349.

[4] Lonni Besançon and Pierre Dragicevic. 2019. The continued prevalence of dichotomous inferences at CHI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[5] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction* 4, 1 (March 1989), 11–44. doi:10.1207/s15327051hci0401_1

[6] Robert Jan Bood, Marijn Nijssen, John Van Der Kamp, and Melvyn Roerdink. 2013. The power of auditory-motor synchronization in sports: enhancing running performance by coupling cadence with the right beats. *PloS one* 8, 8 (2013), e70758.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Stephen A Brewster. 1997. Using non-speech sound to overcome information overload. *Displays* 17, 3 (May 1997), 179–189. doi:10.1016/S0141-9382(96)01034-7

[9] Mark Bubel, Ruiwen Jiang, Christine H. Lee, Wen Shi, and Audrey Tse. 2016. AwareMe: Addressing Fear of Public Speech through Awareness. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 68–73. doi:10.1145/2851581.2890633

[10] João Paulo Cabral and Gerard Bastiaan Remijn. 2019. Auditory icons: Design and physical characteristics. *Applied Ergonomics* 78 (July 2019), 224–239. doi:10.1016/j.apergo.2019.02.008

[11] Maan Isabella Cajita, Christopher E. Kline, Lora E. Burke, Evelyn G. Bigini, and Christopher C. Imes. 2020. Feasible but Not Yet Efficacious: A Scoping Review of Wearable Activity Monitors in Interventions Targeting Physical Activity, Sedentary Behavior, and Sleep. *Current epidemiology reports* 7, 1 (March 2020), 25. doi:10.1007/s40471-020-00229-2 Publisher: NIH Public Access.

[12] Filippo Casamassima, Alberto Ferrari, Bojan Milosevic, Pieter Ginis, Elisabetta Farella, and Laura Rocchi. 2014. A Wearable System for Gait Training in Subjects with Parkinson's Disease. *Sensors (Basel, Switzerland)* 14, 4 (March 2014), 6229–6246. doi:10.3390/s140406229

[13] Michael Cavanagh, Matt Bower, Robyn Moloney, and Naomi Sweller. 2014. The Effect Over Time of a Video-Based Reflection System on Preservice Teachers' Oral Presentations. *Australian Journal of Teacher Education* 39, 6 (June 2014). doi:10.14221/ajte.2014v39n6.3

[14] Francesco Cirillo. 2018. *The Pomodoro technique: The acclaimed time-management system that has transformed how we work.* Currency.

[15] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[16] ELSA Corporation. 2024. ELSASPEAK. https://elsaspeak.com/en/

[17] Nelson Cowan. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences* 24, 1 (2001), 87–114. doi:10.1017/S0140525X01003922

[18] Nelson Cowan. 2008. What are the differences between long-term, short-term, and working memory? *Progress in brain research* 169 (2008), 323–338.

[19] Ádám Csapó and György Wersényi. 2013. Overview of auditory representations in human-machine interfaces. *Comput. Surveys* 46, 2 (Dec. 2013), 19:1–19:23. doi:10.1145/2543581.2543586

[20] Ionut Damian, Chiew Seng (Sean) Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting Social Interactions: Realtime Behavioural Feedback using Social Signal Processing Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 565–574. doi:10.1145/2702123.2702314

[21] Christopher J Darwin, Michael T Turvey, and Robert G Crowder. 1972. An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology* 3, 2 (1972), 255–267.

[22] Prerit Datta, Akbar Siami Namin, and Moitrayee Chatterjee. 2018. A survey of privacy concerns in wearable devices. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 4549–4553.

[23] Julian H. Elliott, Anneliese Synnot, Tari Turner, Mark Simmonds, et al. 2017. Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of Clinical Epidemiology* 91 (Nov. 2017), 23–30. doi:10.1016/j.jclinepi.2017.08.010

[24] Ronald Fischer. 2011. Cross-cultural training effects on cultural essentialism beliefs and cultural intelligence. *International journal of intercultural relations* 35, 6 (2011), 767–775.

[25] Jon Francombe, Russell Mason, Martin Dewhirst, So Bech, et al. 2013. Modelling listener distraction resulting from audio-on-audio interference. In *Proceedings of Meetings on Acoustics*, Vol. 19. AIP Publishing.

[26] Christopher Frauenberger, Tony Stockman, and Marie-Luce Bourguet. 2007. A Survey on Common Practice in Designing Audio in the User Interface. BCS Learning & Development. doi:10.14236/ewic/HCI2007.19

[27] Stavros Garzonis, Simon Jones, Tim Jay, and Eamonn O'Neill. 2009. Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1513–1522. doi:10.1145/1518701.1518932

[28] William W. Gaver. 1986. Auditory Icons: Using Sound in Computer Interfaces. *Human–Computer Interaction* 2, 2 (June 1986), 167–177. doi:10.1207/s15327051hci0202_3

[29] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[30] Sophie S Havighurst, Katherine R Wilson, Ann E Harley, Margot R Prior, and Christiane Kehoe. 2010. Tuning in to Kids: improving emotion socialization practices in parents of preschool children–findings from a community trial. *Journal of Child Psychology and Psychiatry* 51, 12 (2010), 1342–1350.

[31] Michael Hazel, Colleen McMahon, and Nancy Schmidt. 2011. Immediate Feedback: A Means of Reducing Distracting Filler Words during Public Speeches. *Basic Communication Course Annual* 23, 1 (Jan. 2011). https://ecommons.udayton.edu/bcca/vol23/iss1/6

[32] Michael B Himle, Douglas W Woods, John C Piacentini, and John T Walkup. 2006. Brief review of habit reversal training for Tourette syndrome. *Journal of child Neurology* 21, 8 (2006), 719–725.

[33] Victoria Hollis, Alon Pekurovsky, Eunika Wu, and Steve Whittaker. 2018. On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 114 (Sept. 2018), 31 pages. doi:10.1145/3264924

[34] Apple Inc. 2024. User Privacy and Data Use - App Store. https://developer.apple.com/app-store/user-privacy-and-data-use/

[35] Adam Jacks and Katarina L. Haley. 2015. Auditory Masking Effects on Speech Fluency in Apraxia of Speech and Aphasia: Comparison to Altered Auditory Feedback. *Journal of Speech, Language, and Hearing Research : JSLHR* 58, 6 (Dec. 2015), 1670–1686. doi:10.1044/2015_JSLHR-S-14-0277

[36] Nuwan Janaka, Chloe Haigh, Hyeongcheol Kim, Shan Zhang, and Shengdong Zhao. 2022. Paracentral and near-peripheral visualizations: Towards attention-maintaining secondary information presentation on OHMDs during in-person social interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3491102.3502127

[37] Myounghoon Jeon. 2013. Lyricons (Lyrics + Earcons): Designing a New Auditory Cue Combining Speech and Sounds. In *HCI International 2013 - Posters' Extended Abstracts (Communications in Computer and Information Science)*, Constantine Stephanidis (Ed.). Springer, Berlin, Heidelberg, 342–346. doi:10.1007/978-3-642-39473-7_69

[38] Myounghoon Jeon, Pasi Lautala, Chibab Nadri, David N Nelson, Michigan Technological University, and Virginia Polytechnic Institute and State University. 2022. *In-Vehicle Auditory Alerts Literature Review*. Technical Report DOT/FRA/ORD-22/05. https://rosap.ntl.bts.gov/view/dot/60401

[39] Myounghoon Jeon and Bruce N. Walker. 2011. Spindex (Speech Index) Improves Auditory Menu Acceptance and Navigation Performance. *ACM Transactions on Accessible Computing* 3, 3 (April 2011), 10:1–10:26. doi:10.1145/1952383.1952385

[40] Thomas Kosch, Romina Kettner, Markus Funk, and Albrecht Schmidt. 2016. Comparing tactile, auditory, and visual assembly error-feedback for workers with cognitive impairments. In *Proceedings of the 18th international ACM SIGACCESS conference on computers and accessibility*. 53–60.

[41] Frank Krukauskas, Raymond Miltenberger, and Paul Gavoni. 2019. Using auditory feedback to improve striking for mixed martial artists. *Behavioral Interventions* 34, 3 (2019), 419–428. doi:10.1002/bin.1665

[42] Dominika Kwasnicka, Stephan U Dombrowski, Martin White, and Falko Sniehotta. 2016. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. *Health psychology review* 10, 3 (2016), 277–296.

[43] Phillippa Lally, Cornelia HM Van Jaarsveld, Henry WW Potts, and Jane Wardle. 2010. How are habits formed: Modelling habit formation in the real world. *European journal of social psychology* 40, 6 (2010), 998–1009.

[44] Bridget A. Lewis, Jesse L. Eisert, and Carryl L. Baldwin. 2018. Validation of Essential Acoustic Parameters for Highly Urgent In-Vehicle Collision Warnings. *Human Factors* 60, 2 (March 2018), 248–261. doi:10.1177/0018720817742114 Publisher: SAGE Publications Inc.

[45] Guowei Li, Monica Taljaard, Edwin R Van den Heuvel, Mitchell AH Levine, Deborah J Cook, George A Wells, Philip J Devereaux, and Lehana Thabane. 2017. An introduction to multiplicity issues in clinical trials: the what, why, when and how. *International journal of epidemiology* 46, 2 (2017), 746–755.

[46] Ju-Han Lin and Frank Jing-Horng Lu. 2013. Interactive Effects of Visual and Auditory Intervention on Physical Performance and Perceived Effort. *Journal of Sports Science & Medicine* 12, 3 (Sept. 2013), 388–393. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3772579/

[47] Allan MacLean, Richard M. Young, Victoria M. E. Bellotti, and Thomas P. Moran. 1991. Questions, options, and criteria: elements of design space analysis. *Human-Computer Interaction* 6, 3 (Sept. 1991), 201–250. doi:10.1207/s15327051hci0603&4_2

[48] Carolyn Mancuso and Raymond G. Miltenberger. 2016. Using habit reversal to decrease filled pauses in public speaking. *Journal of Applied Behavior Analysis* 49, 1 (2016), 188–192. doi:10.1002/jaba.267

[49] Typing Master. [n. d.]. Speaking Speed Test - Test your speech rate in a minute (WPM). https://www.typingmaster.com/speech-speed-test/.Accessed17July2024

[50] Gerard McAtamney and Caroline Parker. 2006. An examination of the effects of a wearable display on informal face-to-face communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 45–54. doi:10.1145/1124772.1124780

[51] Roisin McNaney, Christopher Bull, Lynne Mackie, Floriane Dahman, Helen Stringer, Dan Richardson, and Daniel Welsh. 2018. StammerApp: Designing a Mobile Application to Support Self-Reflection and Goal Setting for People Who Stammer. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173841

[52] Javier Mendez, Kay Bierzynski, M. P. Cuéllar, and Diego P. Morales. 2022. Edge Intelligence: Concepts, Architectures, Applications, and Future Directions. *ACM Trans. Embed. Comput. Syst.* 21, 5 (Oct. 2022), 48:1–48:41. doi:10.1145/3486674

[53] Pauline Menjot, Lamia Bettahi, Anne-Lise Leclercq, Nancy Durieux, and Angélique Remacle. 2023. Interventions That Target or Affect Voice or Speech Production During Public Speaking: A Scoping Review. *Journal of Voice* (July 2023). doi:10.1016/j.jvoice.2023.06.021

[54] Alaeddine Mihoub and Grégoire Lefebvre. 2017. Social Intelligence Modeling using Wearable Devices. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17)*. Association for Computing Machinery, New York, NY, USA, 331–341. doi:10.1145/3025171.3025195

[55] Christina C. Montes, Megan R. Heinicke, and Danielle M. Geierman. 2019. Awareness training reduces college students' speech disfluencies in public speaking. *Journal of Applied Behavior Analysis* 52, 3 (2019), 746–755. doi:10.1002/jaba.569

[56] Skanda Muralidhar, Jean M R Costa, Laurent Son Nguyen, and Daniel Gatica-Perez. 2016. Dites-moi: wearable feedback on conversational behavior. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*. 261–265.

[57] Elizabeth D. Mynatt. 1994. Designing with auditory icons: how well do we identify auditory cues?. In *Conference Companion on Human Factors in Computing Systems (CHI '94)*. Association for Computing Machinery, New York, NY, USA, 269–270. doi:10.1145/259963.260483

[58] Lutfun Nahar, Riza Sulaiman, and Azizah Jaafar. 2022. An interactive math braille learning application to assist blind students in Bangladesh. *Assistive Technology* 34, 2 (2022), 157–169.

[59] Michael A. Nees and Eliana Liebman. 2023. Auditory Icons, Earcons, Spearcons, and Speech: A Systematic Review and Meta-Analysis of Brief Audio Alerts in Human-Machine Interfaces. *Auditory Perception & Cognition* 0, 0 (June 2023), 1–30. doi:10.1080/25742442.2023.2219201

[60] Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. The preregistration revolution. *Proceedings of the National Academy of Sciences* 115, 11 (March 2018), 2600–2606. doi:10.1073/pnas.1708274114 Publisher: Proceedings of the National Academy of Sciences.

[61] orai.com. 2024. *Orai, an AI-powered app for practicing your presentations.* https://orai.com/

[62] Stephanie M Ortiz, Meghan A Deshais, Raymond G Miltenberger, and Kenneth F Reeve. 2022. Decreasing nervous habits during public speaking: A component analysis of awareness training. *Journal of Applied Behavior Analysis* 55, 1 (2022), 230–248.

[63] Sameer Patil, Roberto Hoyle, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2015. Interrupt Now or Inform Later? Comparing Immediate and Delayed Privacy Feedback. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1415–1418. doi:10.1145/2702123.2702165

[64] S. Camille Peres, Virginia Best, Derek Brock, Barbara Shinn-Cunningham, Christopher Frauenberger, Thomas Hermann, John G. Neuhoff, Louise Valgerður Nickerson, and Tony Stockman. 2008. Chapter 5 - Auditory Interfaces. In *HCI Beyond the GUI*, Philip Kortum (Ed.). Morgan Kaufmann, San Francisco, 147–195. doi:10.1016/B978-0-12-374017-5.00005-5

[65] Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, and Russell Beale. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Transactions on Computer-Human Interaction* 25, 3 (June 2018), 15:1–15:66. doi:10.1145/3196830

[66] Mallory Quinn, Raymond Miltenberger, Takema James, and Aracely Abreu. 2017. An evaluation of auditory feedback for students of dance: Effects of giving and receiving feedback. *Behavioral Interventions* 32, 4 (2017), 370–378. doi:10.1002/bin.1492

[67] Meera Radhakrishnan, Kushaan Misra, and V Ravichandran. 2021. Applying "earable" inertial sensing for real-time head posture detection. In *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events*. IEEE, 176–181.

[68] Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Minh Dinh, Ebrahim Nemati, Jilong Kuang, and Jun Alex Gao. 2022. BreatheBuddy: Tracking Real-time Breathing Exercises for Automated Biofeedback Using Commodity Earbuds. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (Sept. 2022), 213:1–213:18. doi:10.1145/3546748

[69] Pei-Luen Patrick Rau and Jian Zheng. 2019. Modality capacity and appropriateness in multimodal display of complex non-semantic information stream. *International Journal of Human-Computer Studies* 130 (Oct. 2019), 166–178. doi:10.1016/j.ijhcs.2019.06.008

[70] Fred Simard-CEO RE-AK. 2024. Introduction to Smart Glasses — 2024 Tech Review. https://medium.com/antaeus-ar/introduction-to-smart-glasses-2024-tech-review-d871e7d95cb8

[71] Hannah K. Scott, Ankit Jain, and Mark Cogburn. 2024. Behavior Modification. In *StatPearls*. StatPearls Publishing, Treasure Island (FL). http://www.ncbi.nlm.nih.gov/books/NBK459285/

[72] Douglas R Seals and McKinley E Coppock. 2022. We, um, have, like, a problem: excessive use of fillers in scientific speech. 615–620 pages.

[73] GM Siegel and RR Martin. 1967. Verbal punishment of disfluencies during spontaneous speech. *Language and speech* 10, 4 (1967), 244—252. doi:10.1177/002383096701000404

[74] Eric N. Smith, Erik Santoro, Neema Moraveji, Michael Susi, and Alia J. Crum. 2020. Integrating wearables in stress management interventions: Promising evidence from a randomized trial. *International Journal of Stress Management* 27, 2 (2020), 172–182. doi:10.1037/str0000137 Place: US Publisher: Educational Publishing Foundation.

[75] Claire Spieler and Raymond Miltenberger. 2017. Using awareness training to decrease nervous habits during public speaking. *Journal of Applied Behavior Analysis* 50, 1 (2017), 38–47. doi:10.1002/jaba.362

[76] John ER Staddon and Daniel T Cerutti. 2003. Operant conditioning. *Annual review of psychology* 54, 1 (2003), 115–144.

[77] International Standard. 2011. ISO 15006:2011 Road vehicles — Ergonomic aspects of transport information and control systems — Specifications for in-vehicle auditory presentation. https://www.iso.org/standard/55322.html

[78] Tasha R. Stanton and Charles Spence. 2020. The Influence of Auditory Cues on Bodily and Movement Perception. *Frontiers in Psychology* 10 (2020). https://www.frontiersin.org/articles/10.3389/fpsyg.2019.03001

[79] Diane Tam, Karon E. MacLean, Joanna McGrenere, and Katherine J. Kuchenbecker. 2013. The design and field observation of a haptic notification system for timing awareness during oral presentations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, Paris, France. doi:10.1145/2470654.2466223

[80] Felicia Fang-Yi Tan, Ashwin Ram, Chloe Haigh, and Shengdong Zhao. 2023. Mindful Moments: Exploring On-the-go Mindfulness Practice On Smart-glasses. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 476–492. doi:10.1145/3563657.3596030

[81] M. Iftekhar Tanveer, Emy Lin, and Mohammed (Ehsan) Hoque. 2015. Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*. ACM Press, Atlanta, Georgia, USA, 286–295. doi:10.1145/2678025.2701386

[82] Benedict Vorbeck and Christoph Bördlein. 2020. Using auditory feedback in body weight training. *Journal of Applied Behavior Analysis* 53, 4 (2020), 2349–2359.

[83] Bruce N. Walker, Jeffrey Lindsay, Amanda Nance, Yoko Nakano, Dianne K. Palladino, Tilman Dingler, and Myounghoon Jeon. 2013. Spearcons (Speech-Based Earcons) Improve Navigation Performance in Advanced Auditory Menus. *Human Factors* 55, 1 (Feb. 2013), 157–182. doi:10.1177/0018720812450587 Publisher: SAGE Publications Inc.

[84] Xingbo Wang, Haipeng Zeng, Yong Wang, Aoyu Wu, Zhida Sun, Xiaojuan Ma, and Huamin Qu. 2020. VoiceCoach: Interactive Evidence-based Training for Voice Modulation Skills in Public Speaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376726

[85] Ken Wojcikowski and Leslie Kirk. 2013. Immediate detailed feedback to test-enhanced learning: an effective online educational tool. *Medical Teacher* 35, 11 (2013), 915–919.

[86] Ru Zhao, Vivian Li, Hugo Barbosa, Gourab Ghoshal, and Mohammed Ehsan Hoque. 2017. Semi-Automated 8 Collaborative Online Training Module for Improving Communication Skills. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 32 (June 2017), 20 pages. doi:10.1145/3090097

[87] Shengdong Zhao, Felicia Tan, and Katherine Fennedy. 2023. Heads-Up Computing Moving Beyond the Device-Centered Paradigm. *Commun. ACM* 66, 9 (Aug. 2023), 56–63. doi:10.1145/3571722

[88] Chen Zhou, Zihan Yan, Ashwin Ram, Yue Gu, Yan Xiang, Can Liu, Yun Huang, Wei Tsang Ooi, and Shengdong Zhao. 2024. GlassMail: Towards Personalised Wearable Assistant for On-the-Go Email Creation on Smart Glasses. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 372–390. doi:10.1145/3643834.3660683

[89] Zhi Zhou, Xu Chen, En Li, Liekang Zeng, Ke Luo, and Junshan Zhang. 2019. Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. *Proc. IEEE* 107, 8 (Aug. 2019), 1738–1762. doi:10.1109/JPROC.2019.2918951 Conference Name: Proceedings of the IEEE.

# A Methodological Transparency & Reproducibility Appendix (META)

This section describes the supplementary material we shared in the Open Science Framework repository (OSF) at https://osf.io/6vhwn/?view_only=489498d3ac2d4703a17475fc6ca65dfa. The OSF project is currently private and anonymized for blind reviewing, and will become public once the paper is accepted.

## A.1 Preregistration

The preregistration of our evaluation experiment (Sec 6) is available at https://osf.io/yp5xq.

## A.2 Experimental Data

The data for our evaluation experiment is available in our OSF project as a spreadsheet with three tabs:

- *Frequency*. This table contains the normalized frequency of unwanted words for each participant (rows 1 to 24) and each session (columns S1 to S8). The column System specifies the experimental group to which each participant was randomly assigned.
- *Original Data*. This table contains the raw data from which the normalized frequencies were calculated. Again, rows are participants 1 to 24, and for each participant × session combination, the table reports the participant's total speaking time, and the full length of the conversation. The rows are further broken down into individual unwanted words (those identified by the participant in the first session). For each combination of participant × unwanted word × session, the table reports the number of times the unwanted word was uttered and the corresponding frequency (words per minute). Finally, for each participant × session, the table reports the normalized frequency for all unwanted words confounded (column Average).
- *Questionnaire*. This table contains participant responses to the six subjective questions listed in Table 2 and their score to the Nasa-TLX questionnaire.

## A.3 Interview Materials

In our OSF project, we share the interview questions used in the training and post-training phases.

## A.4 *WSCoach* Software

In our OSF project, we share the python source code for the WSCoach software we used in our evaluation. We ran this software on a desktop computer with a 3080Ti Graph Card connected to a Huawei Eyewear via Bluetooth, but any audio mic (e.g., Bluetooth earbuds) can be used as the audio source and laptops with graphic cards can run the program. The computer should have Python3 installed. Further instructions are available in the README file.

## A.5 Additional Information About *Orai*

Figure 9 (also available in the OSF project) shows screenshots of the *Orai* application's user interface, enabling diverse post-conversation analyses. For more information please visit https://orai.com/.

## A.6 Full Analysis of Secondary Measurements

Here we report the full analysis of secondary measurements, which we only partially reported in Sec 7.3 for space reasons.

- *Confidence in Speaking*. The mean response for this question was 4.8, 95% CI [4.0, 5.5] for *WSCoach* and 4.5, 95% CI [3.6, 5.4] for *Orai*, see Figure 8 (a). The $U$-statistic for the difference is $U = 71$, $p = .99$. Therefore, we have no evidence that *WSCoach* and *Orai* differed in terms of Confidence in Speaking.

- *Awareness of Speaking Unwanted Words.* The mean score was 5.8, 95% CI [5.2, 6.4] for *WSCoach* and 5.5, 95% CI [5.2, 5.8] for *Orai*, see Figure 8 (b). The $U$-statistic for the difference is $U = 54$, $p = .27$. Therefore, we have no evidence that *WSCoach* and *Orai* differed in terms of Awareness of Speaking Unwanted Words.
- *Self-Ranked Conversation Quality.* The mean score was 5.4, 95% CI [4.9, 6.0] for *WSCoach* and 5.1, 95% CI [4.5, 5.7] for *Orai*, $U = 54$, $p = .31$, see Figure 8 (c). Again we have no evidence for a difference on this metric.
- *Distraction of Feedback.* The mean score was 4.1, 95% CI [3.0, 5.2] for *WSCoach* and 1.8, 95% CI [1.1, 2.4] for *Orai*, $U = 117$, $p = .003$, see Figure 8 (d). Therefore, we have good evidence that *WSCoach* generated more distraction on average than *Orai* during conversations.
- *NASA TLX for Being Aware of Unwanted Words.* The mean score was 41, 95% CI [25, 57] for *WSCoach* and 30.64, 95% CI [20, 41] for *Orai*, $U = 55$, $p = .35$, see Figure 8 (e). Again, we have no evidence of a difference here either.
- *Confidence in Conversation (post-training phase).* The mean score was 5.3, 95% CI [4.9, 5.6] for *WSCoach* and 4.3, 95% CI [3.8, 4.9] for *Orai*, $U = 28$, $p = .01$, see Figure 8 (f). Therefore, we have evidence that *WSCoach* made people feel more confident in their conversation than *Orai* after the training was over.
- *Awareness of Avoiding Unwanted Words (post-training phase).* The mean response was 5.4, 95% CI [4.9, 5.9] for *WSCoach* and 4.8, 95% CI [4.0, 5.6] for *Orai*, see Figure 8 (g). The $U$-statistic is $U = 53$, $p = .25$. Therefore, we have no evidence that *WSCoach* and *Orai* differed in terms of Awareness of Avoiding Unwanted Words.

## A.7 Auditory feedback audio samples

Our OSF project contains audio samples we used in our pilot studies. There is one example for each of the five feedback techniques we used: Earcon, Spearcon, Spindex, Speech, and Lyricon. We also provide an example of Auditory Icon.

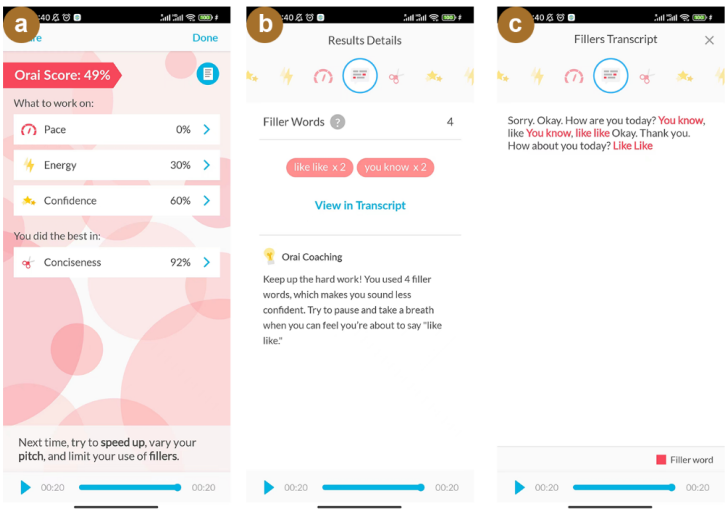


Fig. 9. The *Orai* application. (a) *Orai* provides feedback on key communication metrics. (b) *Orai* Lists the count of filler words or unwanted words after conversation visually. (c) Users can see the highlighted filler words in audio transcription and play the recording.