



Citation for published version:

Röddiger, T, Clarke, C, Wolfram, D, Budde, M & Beigl, M 2021, EarRumble: Discreet Hands- and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction. in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.*, 743, CHI Conference Proceedings, Association for Computing Machinery, pp. 1-14. <https://doi.org/10.1145/3411764.3445205>

DOI:

[10.1145/3411764.3445205](https://doi.org/10.1145/3411764.3445205)

Publication date:

2021

Document Version

Peer reviewed version

[Link to publication](#)

© ACM, 2021. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, {May 2021, Article no. 743} <http://doi.acm.org/10.1145/3411764.3445205>

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

EarRumble: Discreet Hands- and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction

Tobias Röddiger
roeddiger@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Christopher Clarke
chris.clarke@lancaster.ac.uk
Lancaster University
Lancaster, United Kingdom

Daniel Wolffram
wolffram@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Matthias Budde
budde@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Beigl
beigl@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

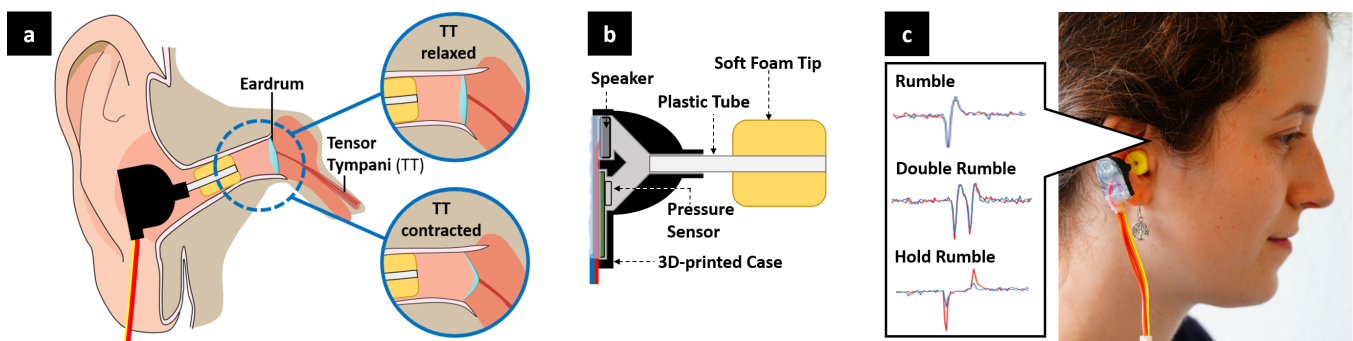


Figure 1: We present EarRumble, a technique that uses “ear rumbling” for interaction. (a) The tensor tympani muscle can be contracted voluntarily which displaces the eardrum and induces a pressure change within the sealed ear canal; (b) Custom-built earables detect ear rumbling using an in-ear pressure sensor; (c) Eyes- and hands-free discreet input can be provided by performing different rumbling gestures by voluntarily contracting the tensor tympani muscle.

ABSTRACT

We explore how discreet input can be provided using the tensor tympani - a small muscle in the middle ear that some people can voluntarily contract to induce a dull rumbling sound. We investigate the prevalence and ability to control the muscle through an online questionnaire (N=192) in which 43.2% of respondents reported the ability to “ear rumble”. Data collected from participants (N=16) shows how in-ear barometry can be used to detect voluntary tensor tympani contraction in the sealed ear canal. This data was used to train a classifier based on three simple ear rumble “gestures” which achieved 95% accuracy. Finally, we evaluate the use of ear rumbling for interaction, grounded in three manual, dual-task application scenarios (N=8). This highlights the applicability of EarRumble as a low-effort and discreet eyes- and hands-free interaction technique that users found “magical” and “almost telepathic”.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Gestural input**; **Interaction devices**.

KEYWORDS

tensor tympani muscle, discreet interaction, subtle gestures, earables, hearables, in-ear barometry

ACM Reference Format:

Tobias Röddiger, Christopher Clarke, Daniel Wolffram, Matthias Budde, and Michael Beigl. 2021. EarRumble: Discreet Hands- and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3411764.3445205>

1 INTRODUCTION

Earphones are widely adopted by consumers because they provide private audio channels to listen to music, podcasts, or audiobooks, to watch the latest movies and TV series whilst commuting, or to make hands-free phone calls. Earables are smart earphones with on-board sensors and real-time data capabilities which present new opportunities for interaction with mobile devices - where interactions may occur when the user is pre-occupied, or in public spaces surrounded by other people. Input techniques using “subtle” or “motionless” input gestures are desirable in mobile contexts

because they take into consideration the social context of mobile device usage [14].

Input techniques with little to no movement avoid the inconvenience of techniques requiring large physical effort (e.g. hand gestures) [22] and are more socially acceptable to spectators [46]. The latter benefit also has the advantage from a user’s perspective of maintaining a level of privacy over the interaction to avoid unwanted attention [42]. Microgestures and hands-free input approaches enable users to interact with their device without disrupting other tasks they may be performing, for example manual tasks such as writing a letter or driving a car [55]. Earables can be used to detect input by tapping the earphone itself [9, 34], or as a sensing platform to detect more advanced gestures performed directly on [28], around [31, 56], or in front of [39] the ear. The earable platform also presents unique opportunities for interaction, such as the use of in-ear barometry to detect head gestures [3]. However, the use of such interactions, in addition to others such as opening and closing the mouth, raises social acceptability issues [46] and there is large scope for the exploration of more discreet methods of input using the earable medium.

We introduce *EarRumble*, an interaction technique which uses contraction of the tensor tympani muscle inside the middle ear, see Figure 1(a). The tensor tympani muscle is the second smallest muscle found in the human body, and is used for dampening loud noises [26]. Interestingly, a subset of the population has voluntary control over the contraction of the tensor tympani muscle which causes in-ear vibrations when contracted [43, 50]. As the tensor tympani contracts it tightens the eardrum and the volume encapsulated in the ear canal rises, resulting in a reduction in pressure. *EarRumble* measures pressure changes within the sealed ear canal to unobtrusively detect contraction of the tensor tympani [24], see Figure 1(b).

We explore how we can exploit this phenomenon, also known as ear rumbling, for interaction with mobile devices using simple, discreet gestures - however, not everyone has voluntary control of the tensor tympani muscle. To inform how many people could hope to use ear rumbling for interaction, and to demonstrate that rumbling is a viable opportunity for others to pursue, we begin by investigating the prevalence of rumbling through an online questionnaire (N=198) which revealed that 43.2% of respondents had the ability to voluntarily contract the tensor tympani. As contraction of the muscle is unlikely to occur voluntarily on a regular basis, it is unclear what level of control users have over the muscle for the purposes of interaction. To address this, we provide initial insights into the complexity of the interaction design space through analysis of different ear rumble gestures (e.g. sequential rumbles, holding the rumble) and gather feedback on user perception of how easy and comfortable they are to perform on demand to determine their viability. We collect data from participants with the ability to ear rumble (N=16), and analyse the characteristics of three different ear rumble gestures, see Figure 1(c). Using the data collected we develop a recognition pipeline which detects ear rumble gestures from everyday activities that may also induce pressure changes in the ear canal, with up to 95% accuracy with real-time performance. Finally, we explore how *EarRumble* can be used for interaction in two exemplar applications (receiving phone calls and an audio player) using three manual, dual task application scenarios that one

might face when using earables in everyday life. Users (N=8) provided positive feedback, describing how the *EarRumble* technique felt “magical” and “telepathic”, and highlighting how *EarRumble* is a low-effort, hands- and eyes-free input technique.

In sum, our paper provides the following contributions: (i) *EarRumble*, a hands- and eyes-free, discreet input technique based on voluntary control of the tensor tympani muscle found in the middle-ear, and sensed through in-ear barometry; (ii) An indication of how prevalent ear rumbling is, how easy it is to perform on demand, and how comfortable it is through an online questionnaire; (iii) Data-driven insights into how well users can contract the tensor tympani muscle in the context of interaction; and (iv) Insights into how *EarRumble* can be used for interaction, grounded in real-world applications involving dual task scenarios.

2 BACKGROUND AND RELATED WORK

We first introduce subtle and discreet interaction in HCI research. Then, we summarise relevant publications of ear interactions and sensing, and finally, we provide the medical background of the tensor tympani muscle.

2.1 Subtle and Discreet Interactions

This paper introduces a subtle and discreet interaction technique that requires low effort and can be hidden from others – two areas that were recently highlighted as part of a systematic investigation into subtle interaction in the HCI literature [42]. Motivations for *doing less* in interaction include (a) to make interaction smaller and more comfortable [32] so that they do not cause physical discomfort [16], (b) being always available [32] and/or (c) “to execute a secondary task, for example controlling mobile applications, without interrupting the manual primary task, for instance, driving a car” [55]. Costanza et al. promote the term *intimate* interfaces, meaning subtle, discreet and unobtrusive control of mobile devices [14]. Systems that enable subtle interaction, but involve technology that itself is not subtle include *Gunslinger* (two 3D cameras for barehand gestures) [32], the *Magic Ring* (finger-worn wearable with accelerometer to detect small finger gestures) [25] and *Bitey* (teeth clicking through a head-mounted bone microphone) [6]. We contrast this in our approach by integrating pressure sensors with commodity earphones, and the act of contracting the tensor tympani does not necessarily require an externally noticeable user gesture or facial expression. Other discreet interaction techniques of note are *Itchy Nose* by Lee et al. which employed EOG sensors embedded in the frame of smart glasses to detect small finger gestures performed on the nose [30], and Gallego Cascón et al.’s *ChewIt* – an intraoral input device that resembles an edible object and allows performing various hands-free input-operations [19].

Another common motivation for subtlety is to enable socially acceptable interaction, meaning to not “*disrupt [...] others in the vicinity, or others in the group*” [45]. Users may desire privacy, e.g. to protect private texts, passwords, or PIN entry. Taking this to extremes can mean completely hiding the fact that interaction happens at all. There are also application-specific motivations for subtle interaction: many researchers have approached discreet interaction in the context of different modalities, ranging from micro-gestures with the hand [18], fingers [11], gaze [15, 27], and oral interfaces

[19, 52]. The use of the ear is underexplored, and EarRumble enables interaction in which users can *do less* and are not impaired in other actions, as well as provide *hidden* interaction that is undetectable by, and non-disruptive to, others.

2.2 Ear Interaction and Sensing

Related work has explored a plethora of earable interaction and sensing techniques for human-centered applications. Similar to our paper, in-ear barometry was applied to classify tongue [51] and face gestures [3]. EarRumble expands upon the sensing principle’s design space with a set of three new discreet ear-based gestures. Existing, hand-based interaction techniques detect tap gestures on [9, 34, 56], mid-air interactions in front of [31, 39], and touch gestures around [31, 56] the ear. Audio-based principles pointing in the ear were applied to detect changes in facial expression [1], to classify eating episodes [2], and for authentication [5]. Electric-field and proximity sensing were also used to detect movement of the face [7, 36, 47]. Interactions making direct use of the ear rely on deforming and pulling the ear [28, 31], use ear wiggle as a subtle input technique [35], or use the ear as touch input to assist the visually impaired [54]. None of the work that we identified regarding interactions on, around, or by the ear investigated voluntary control of the tensor tympani muscle as an active input mechanism.

2.3 Tensor Tympani Muscle

The tensor tympani muscle sits in the human middle ear and actuates the tympanic membrane during the middle ear reflex [41, 50]. The tensor’s subconscious contraction accompanies vocalization and swallowing [29], or expecting a startling sound [17], and the ability to voluntarily contract the tensor tympani muscle has been discussed for over one hundred years [44]. Due to the vibrations induced by the tensing muscle, a rumbling sound can be heard during contraction [44], and those who can voluntarily contract the muscle often describe it as flexing, activating, or moving a muscle inside their ear¹. The sound might be imitated by firmly tensing one’s fist and pressing it on the ear which creates a comparable dull rumbling sound. Little information is available about the prevalence of the ability to consciously control the tensor tympani muscle [4]. In this paper, we conduct an online questionnaire to provide insights into the prevalence of voluntary contraction, and explore the applicability and constraints of an input technique that relies on voluntary tensor tympani control.

In the past, the medical field documented multiple principles to sense tensor tympani muscle contraction [41]. Electromyography is the most invasive method as it requires surgery for the placement of electrodes on the tensor [48]. Otologists commonly measure the acoustic impedance of the ear which might also be used to detect tensor contraction [40, 41]. However, the technique requires playing an 800 Hz probing tone which might create discomfort. It is also possible to use a camera otoscope to detect contraction. This is used in section 5 for validation that participants are contracting the tensor tympani by visually observing the eardrum displacement as a result of tensor tympani contraction. However, we opted against camera-based technology for our implementation of EarRumble because of power consumption, computational complexity, occlusion

(e.g. ear wax), and the need to focus the camera on the eardrum. Alternatively, in-ear barometry can be used to measure the displacement of the eardrum during contraction through pressure changes in the sealed ear canal [24]. We use this approach in EarRumble because it does not require playing a constant tone, is cheap to realise with off-the-shelf components, and can be incorporated into a pair of earphones. Previous work has also demonstrated the utility of in-ear barometry for interaction through head gesture and facial expression detection [3].

3 EARRUMBLE

We introduce *EarRumble*, an interaction technique based on the voluntary contraction of the tensor tympani muscle found in the human middle ear - a phenomenon also known as ear rumbling. The contraction of the tensor tympani is detected using in-ear barometry, i.e. measurement of pressure changes within the ear canal, using a custom-built earable consisting of commercial off-the-shelf (COTS) components and a custom 3D-printed enclosure, see Figure 2. A thresholding detection algorithm and feature-based machine learning classifier are applied to recognise three basic ear rumbling gestures from the raw pressure signals, see Figure 3.

3.1 Concept

Figure 1(a) illustrates the underlying principle of the EarRumble interaction technique. Upon contraction of the tensor tympani muscle, the eardrum displaces inward. As the soft foam earcaps worn by the user seal the ear canal, the volume increases while the encapsulated air remains constant which results in a pressure drop. After the relaxation of the muscle, the eardrum returns to its original position. This leads to a pressure wave that is pushed outwards of the ear canal to produce a positive pressure peak, as shown in the *Rumble* signal in Figure 1(c). As the sealing of the ear canal is not perfectly air-tight, the pressure equalizes over time. Therefore, holding the tensor tympani contracted does not yield a constant low pressure reading, however releasing the muscle still produces a sufficiently pronounced pressure peak in the opposite direction, as shown in the *Hold Rumble* signal in Figure 1(c). We utilise the changes of pressure to derive contraction events of the muscle which can be measured with a standard, off-the-shelf pressure sensor. We assume the tensor tympani can be in one of two states – relaxed or contracted – and we focus our exploration on the use of three basic ear rumble gestures based on insights from sections 4 and 5:

- Single rumble – a quick contraction of the tensor tympani
- Double rumble – two contractions of the tensor tympani in quick succession
- Hold rumble – contraction of the tensor tympani for approximately one second

3.2 Hardware

Figure 2 illustrates the assembly of our custom-built device. To realise in-ear barometry, we use the Bosch BME280 pressure sensor (2.5 x 2.5 x 0.93 mm) sampling at 32 Hz. We removed the speaker from a pair of commercially available earbuds (Sony MDR-EX110LP) and 3D-printed a custom earplug case that encapsulates the different components in a single device. The case splits the enclosed air

¹Ear Rumlbers Assemble subreddit: www.reddit.com/r/earrumlbersassemble/

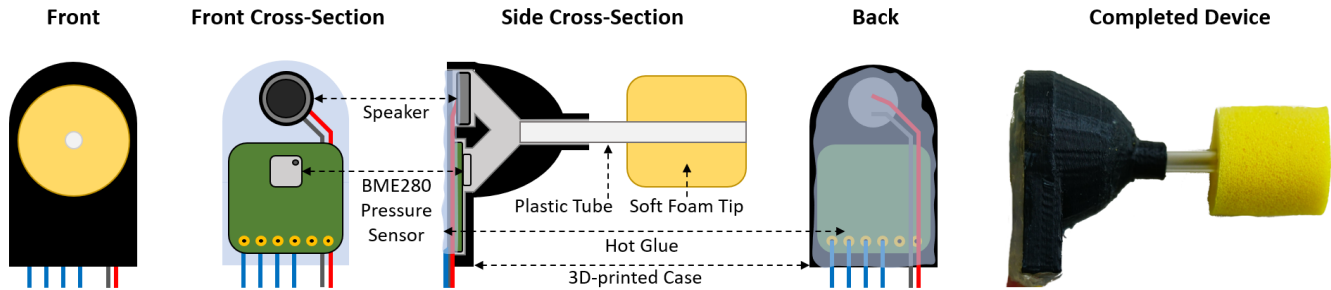


Figure 2: The custom-built device used to realise the EarRumble technique. Each earbud contains a speaker, which provides the usual audio capabilities of a pair of headphones, and a BME280 pressure sensor which is used to detect the ear rumble gestures by measuring the changes of pressure inside the ear canal. The soft foam tip and hot glue provides an air-tight seal within the ear canal. A custom 3D-printed enclosure houses the components and provides two separate channels for the speaker and pressure sensor.

into two channels, one directed towards the speaker and the other towards the pressure sensor. This was done to minimise the volume of air enclosed within the ear canal, and maximise the change in pressure. To ensure tight sealing of the ear canal, we use foam earplugs (Etymotic Research Disposable eartip ER1-14A, 13mm diameter). Before insertion users firmly squeeze the tip, which then expands within the ear to create a tight seal. To increase airtightness further and to keep the electronic components in place, we seal the 3D-printed case by applying hot glue on the backside of the components. The manufacturing process is the same for left and right earbuds, except that the device uses stereo sound which plays the respective channel on either of the ears. The pressure sensors in the earphones connect to an ESP32 MCU breakout board using I2C and data is transferred to a PC using serial communication. The audio signal connects to the same workstation using the aux connector.

3.3 Recognition Pipeline

Figure 3 illustrates the final recognition pipeline of EarRumble. Other classifiers and their performance are evaluated in section 5.

3.3.1 Detection. Activity detection uses a 360 ms sliding window with step size of 120 ms to decide if a window of the pressure signal contains activity. The detector computes the sum of absolute difference (SAD) within each window, and those with an SAD above 20 Pascal (PA) are flagged as containing activity. The threshold value was deduced from different rumble activity and non-activity samples that we collected during section 5, and ensures that > 95% of all samples are correctly detected. We define a correct detection as one in which more than 75% of the gesture was detected as activity. An activity is considered complete and passed to the classifier after four consecutive windows without activity.

3.3.2 Classification. Once an activity is detected, it is passed to the classification stage to determine whether an ear rumble has been detected. Features are extracted from the signal, before being passed to a gesture classifier. We evaluated four different classifiers: a radial basis function (RBF) kernel support vector machine

(SVM), k-nearest neighbours (kNN), random forest (RF), and gradient boosting (XGBoost). The XGBoost classifier yields the highest overall accuracy at 0.95. More details are presented in section 5.

For classification we selected features that we self-defined, from the work by Ando et al. [3], in addition to systematically derived features from tsfresh [12] using the data gathered in section 5. All features are computed on the zero mean-shifted signal to account for drift in the pressure readings. To reduce the initial collection of 1,618 features systematically, we apply importance selection with XGBoost feature importance scores. If a feature was deemed relevant for the left or right ear only, we compute it for both ears to account for different laterality conditions. As a result, we used a set of 70 features (35 per ear) for feature extraction, which are then passed to the classifier. The average computation time of all features across all collected samples was 46.2 ms (Intel Core i7-9700KF 8 x 3.7 GHz).

For replicability, the features we used include the number of, and the mean distance between peaks (both negative and positive), the absolute difference of the first two peaks, and the minimum and maximum values of the signal and also their locations. Additionally, the absolute difference and ratio of maximum to mean, and also the minimum and maximum slope and intercept over a five sample rolling window. We use the sum of absolute differences (SAD) of the whole signal and also SAD of four even sequences that we cut each sample into. Finally, auto-correlation of the signal (lag: 2, 5, 9, 32), the variance of quantile changes (ql - qh: 0.2 - 1.0, 0.0 - 0.8), the spectral welch density (c: 2, 5), the continuous wavelet transform of the Ricker wavelet (c / w: 0 / 20, 2 / 2, 4 / 5, 8 / 20) and the FFT coefficients of the signal (c: 1, 2).

4 PREVALENCE OF TENSOR TYMPANI MUSCLE CONTROL

Although not part of the acoustic reflex, the involuntary contraction of the tensor tympani muscle (tympani reflex) helps prevent ear damage from loud noises [26]. However, not everyone can voluntarily contract the muscle to cause in-ear vibrations, and there is currently no data about the prevalence of ear rumbling, nor how well people can voluntarily control the tensor tympani muscle. In

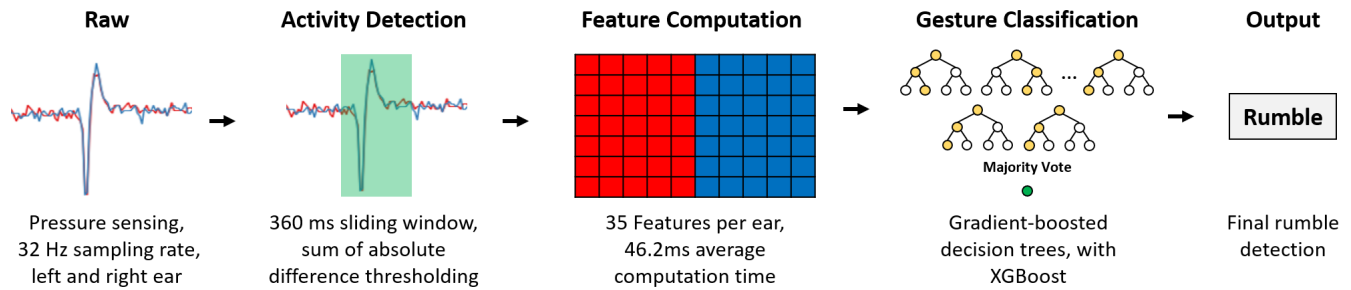


Figure 3: The EarRumble recognition pipeline – the raw pressure signals are sampled at 32 Hz from both ears. A sliding window of 360 ms is used to detect if activity is present in the signal. 35 features are sampled from each ear, which are then passed to a gesture classifier to identify if the signal contains an ear rumble.

this section, we survey a large sample of participants remotely using an online questionnaire, predicated on the basis that voluntary contraction of the tensor tympani can be self-reported due to audio feedback during the contraction of the muscle.

We look to gain insight into what proportion of the population can voluntarily control the tensor tympani muscle - to inform how many people could hope to use ear rumbling for interaction. Of those who can voluntarily control contraction of the tensor tympani, we sought to gain deeper insight with respect to the level of control, isolation, and laterality when performing the ear rumbling. We queried the level of control participants have when contracting the muscle to inform the potential complexity of the interaction design space. We investigate the level of discreteness afforded by ear rumbling as an interaction technique, which is dependent upon whether the tensor tympani can be contracted in isolation of other physical movements or facial gestures. Finally, we seek to discover whether participants can perform ear rumbling in one ear or both, which informs whether we can treat signals from the ears independently or in combination.

4.1 Design and Procedure

To minimise non-response bias – where those who can not ear rumble find the questionnaire less appealing – we recruited participants using neutral, context-free online ads and a social media post (Twitter) that did not reveal information on the nature of the study. Participants could use any device with browsing capabilities to fill in the survey, and no reward was offered for participating in the study. Firstly, participants were presented with the information sheet and relevant consent forms. Participants were not admitted to the study if they self-reported acute ear-related health conditions or were not at least 18 years of age.

The first page introduced the concept of “ear rumbling” by describing the contraction of the tensor tympani muscle. This was illustrated using an animated image of the human ear and a short textual description based on observations from people who have voluntary control, related work reporting the phenomenon, and by talking to an ear, nose, and throat doctor. Participants were then instructed to move to a quiet environment and to remove headphones, with explicit instructions for those who wear hearing aids to leave them on. We then asked participants to attempt to activate, flex, or move the muscles inside their ear to produce a rumbling

sound. In addition, we added clarification that this is not to be confused with ear wiggling, and that some people can only perform ear rumbling when performing other actions (e.g. when yawning, swallowing, with closed eyes, or with the mouth open). Participants were then asked whether they could hear a “rumble or vibration”, or a “clicking, crackling or popping” sound, or both. We asked about the clicking, crackling, or popping sound as such sounds might be induced by opening the Eustachian tube for pressure exchange [38], rather than contraction of the tensor tympani muscle.

Participants who reported that they can perform ear rumbling, either with or without crackling noises in addition, were asked which ear they could hear the rumbling in and whether they could perform the rumbling independently of other actions, such as closing the eyes, blinking, or swallowing. They were also asked to complete two 7-point Likert items (1: strongly disagree, 7: strongly agree), one asking whether the rumbling is easy to perform on demand, the other asking whether it is comfortable to perform. To investigate the level of control users have of contracting the muscle, participants were asked whether they could perform the rumbling in quick succession (e.g. one ear rumble directly after another), and whether or not they could control the duration of the rumbling (e.g. hold an ear rumble for one second). The order in which these questions were presented was counterbalanced in the event that performing one movement made it more difficult to perform the other. If participants answered yes to either of these, they were asked the same two Likert items regarding ease of performance on demand and comfort, in addition to whether any additional action was required for the rumbling. The whole survey took around seven minutes to complete.

To validate the results from the online questionnaire, the same questionnaire was given to sixteen participants who completed the questionnaire as part of an in-person evaluation in section 5. These responses were validated visually using a USB otoscope and form a separate dataset. We compare the responses to the Likert items from the online participants with those from section 5 to see if any statistically significant differences exist between the two datasets.

4.2 Results

208 participants completed the study, from a total of 1,399 clicks on the adverts. After data cleaning, there were 192 completed data sets (110 male, 78 female, 1 other, 4 preferred not to answer, age:

$M = 40.1$, $SD = 13.7$, $min = 18$, $max = 76$). Eight participants had reduced hearing abilities (4 medical and 4 self-diagnosed), three wore hearing aids on both ears, and one participant was deaf.

4.2.1 Prevalence. Out of 192 participants, 83 (43.2%) reported that they could produce a rumbling or vibrating sound on at least one ear. Using the normal approximation interval, this results in a 95% confidence interval between 36.2% to 50.2%. Of the 83 who reported that they could produce a rumbling, 18 reported that they heard a crackling, clicking or popping sound in addition to the rumbling. Out of the 18 participants who reported rumbling and popping sounds, 9 said that they can do the rumbling sound independent of the crackling. In addition to those who reported some form of rumbling, 44 participants reported a crackling sound, but no rumbling. Those who did not report a rumbling sound were not asked to complete the remainder of the study. The deaf participant did not report rumbling, however participants with reduced or no hearing ability might still be able to contract the tensor tympani, but are lacking the audible feedback loop.

4.2.2 Laterality. Of those who could rumble, 68 participants (81.9%) reported the rumbling sound on both ears simultaneously, 11 (13.3%) in isolation on the left, and 19 (22.9%) in isolation on the right ear. Overall, 14 (7.1%) participants reported that they could perform rumbling on both ears and also in isolation on one ear - suggesting a high level of control of the muscle contraction. No participants reported the ability to perform rumbling on both ears and in isolation on both sides.

4.2.3 Control. Out of all those who could rumble, 74 (89.2%) reported the ability to perform rumbles in quick succession, and 71 participants (85.5%) could change the duration of rumbles. 67 participants reported that they could both perform rumbles in quick succession and change the duration of the rumbles. Interestingly, three of the participants with reduced hearing abilities could perform all rumbling variations (1 medical, 2 self-diagnosed). Figure 4(a)-(c) shows that participants perceived ear rumbling to be easy to perform on-demand and comfortable. We performed a Friedman test of those who reported that they could perform all types of ear rumble, which revealed a significant difference for responses about how easy ear rumbling is to perform on-demand, but no difference for the reported level of comfort across the rumble variations. Posthoc Wilcoxon rank-sign tests revealed that performing a single ear rumble was perceived to be much easier to perform on demand compared with both successive rumbles ($Z=37.5$, $p < .01$), and when holding a rumble for a prolonged duration ($Z=27.5$, $p < .01$). There was no statistically significant difference between successive rumbling or holding for a prolonged duration.

4.2.4 Isolation. The majority of participants did not have to perform secondary actions when contracting the tensor tympani muscle. Five participants required a secondary action when performing repetitive rumbling, and six participants when holding the rumbles for a longer duration. Figure 4(d) shows the breakdown of the three types of secondary action which were required by these participants: closing eyes, blinking, and yawning.

4.2.5 Comparison to Lab Dataset. To validate the reliability of the Likert items administered online, we compared the responses

from the online participants with responses from participants who completed the same questionnaire in-person as part of the data collection performed in section 5. Mann-Whitney U tests showed that no significant differences existed between the responses across the three types of rumbling (single, changing duration, and successive rumbling) for either of the Likert items. Both data sets showed that participants found ear rumbling easy to perform on-demand, and comfortable, see Figure 4.

4.2.6 Qualitative Results. Five participants commented that they remember being able to perform ear rumbling since they were a child (P492, P556, P1320, P1809, P1709) and were surprised to learn that not everyone has that ability. Another participant said that they can recall involuntary rumbling during stressful situations (P1320). One participant expressed the urge to keep rumbling after rumbling once (P480). Another participant with mild tinnitus mentioned that the rumbling can only be perceived when the surrounding area is very quiet. One participant expressed that rumbling led to a tense feeling in the cheek muscles.

4.3 Discussion

These results demonstrate that a substantial proportion of those who responded had the ability to voluntarily contract the tensor tympani muscle, and generally participants reported that ear rumbling is easy to perform on demand and comfortable. It is unclear if and how the muscle contraction might be learnable for those who do not possess the ability to voluntarily contract the tensor tympani, or indeed if the muscle can be strengthened over time for those who can. Very few people required secondary actions (such as closing the eyes) to induce the ear rumble, which makes it ideal as a discreet interaction technique that is externally hard to notice. Of those who reported they could perform an ear rumble, 80.7% reported that they could change the duration of the ear rumble, and perform multiple ear rumbles in succession. Using this knowledge, we can investigate how a simple gesture set that leverages these characteristics can provide simple input capabilities using contraction of the tensor tympani. We explore this in more depth in the next section.

The results reported in this section are predicated on the ability of participants to accurately self-report the ability to voluntarily contract the tensor tympani. In section 5, all participants ($N=18$) were initially recruited based on self-reporting the ability to ear rumble, which was validated using a camera otoscope and confirmed in all cases. We also note that the subjective responses of the in-person participants were not significantly different compared with the online survey. However, due to the nature of the online survey we are unable to physiologically validate the ability to voluntarily contract the tensor tympani for the remote participants, and it is important to note that the data collected does not stand as physiological evidence.

5 DATA COLLECTION AND ANALYSIS

We conducted a study to investigate the feasibility of detecting ear rumbling across a range of participants using in-ear barometry. We seek to understand how well users can perform ear rumbling on-demand, how comfortable they find it, and the general characteristics when performing ear rumbles to inform interaction design.

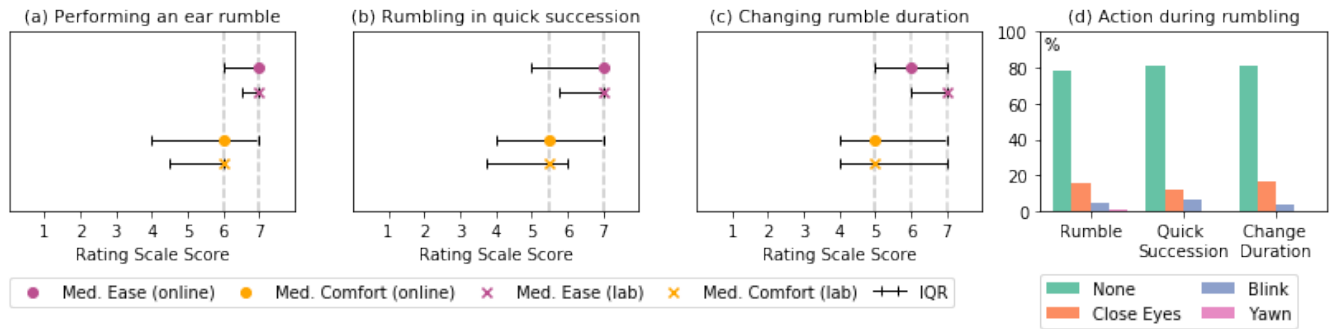


Figure 4: The median and inter-quartile range (IQR) for online dataset from section 4 and lab participants from section 5 for the two Likert items – ease to perform on demand and comfort for performing – for (a) a rumble, (b) rumbling in quick succession, and (c) changing the duration of rumbles (1: strongly disagree SD, 7: strongly agree SA). Figure (d) shows the fractions of required secondary movement during different rumbling variations.

Based on initial data exploration, we focus the data collection on the use of three ear rumble gestures: single, double, and hold rumble.

We also investigated participants’ ability to perform repetitive sequential rumbling, where they were asked to contract the tensor tympani multiple times in quick succession. We wanted to understand the level of control participants had over the muscle contraction, because a higher level of control opens up the opportunities for the use of switch scanning interfaces [49], rhythmic patterns [20], or beat synchronization [53, 57] for interaction.

Previous work has demonstrated that tensor tympani contraction might occur during swallowing and vocalization [29], and that in-ear pressure sensing can be used to detect a number of different gestures that users may naturally perform in their everyday lives, such as opening and closing their mouth [3]. This begs the question as to whether we can accurately distinguish pressure changes from ear rumbling compared with similar everyday activities. With this in mind, we also capture data on the following actions which may induce pressure changes in the ear canal: opening and closing the mouth, reading out loud, drinking and swallowing, and chewing gum. The tensor tympani muscle contraction may also be induced by sound [17], therefore we investigate ear rumbling under two conditions: in silence, and with music playing through the earphones.

5.1 Participants and Apparatus

We recruited 18 participants through e-mail and a university Facebook group. Prior to the study, participants self-reported that they could perform ear rumbling, which was validated using an otoscope and confirmed in all cases. The study was conducted according to national COVID-19 regulations and within the university’s safety guidelines. Participants wore the custom-built, in-ear pressure sensing device on both ears, see subsection 3.2. Each participant’s outer ear canal was measured using a caliper, and two participants were excluded because of insufficient sealing of the ear canal with the ear buds (> 14 mm external ear canal diameter). The final dataset consists of 16 participants (13 male, 3 female, Age: $M = 24.7$ $SD = 2.63$, Ear canal width: $M = 8.0$ mm $SD = 1.6$ mm , Ear canal height:

$M = 11.2$ mm $SD = 1.8$ mm). All participants reported that they had no hearing loss, and none of the participants wore a hearing aid.

5.2 Design and Procedure

Participants began the study by completing the same questionnaire about their ability to ear rumble featured in section 4, and all participants reported that they had not participated in the online questionnaire in section 4. Following this, the researcher measured the ear canal and verified that the ear rumbling reported by participants was caused by contraction of the tensor tympani muscle. This was validated by visual inspection of the eardrum with a Teslong USB digital in-ear camera otoscope. Participants were then asked to wear the EarRumble earphones and ensure a tight fit so that pressure differences could be detected.

Participants performed nine activities – four ear rumbling gestures (single, double, hold, repetitive), four everyday activities (opening/closing mouth, reading out loud, drinking, and chewing gum), in addition to an activity where users were asked to do nothing. A display was used to indicate which activity the participant should perform. Each activity was preceded by a five second on-screen countdown, after which participants were instructed to execute the activity in a five second window. We asked participants to perform the rumbling gestures immediately and as quickly as possible after the countdown. Data was recorded from the start of the countdown, until the five seconds of the activity had elapsed. After each activity there was a five second break. After performing one of the rumbling activities, participants completed two 7-point Likert items (1: strongly disagree to 7: strongly agree):

- *Ease-on-demand*: The [rumbling activity] was easy to perform on demand.
- *Comfort*: The [rumbling activity] was comfortable to perform.

Our study follows a within-subject design, where all participants performed all activities. Participants performed the activities in two blocks: one in silence, and one with music playing (Symphony No. 5 by Ludwig van Beethoven). The order of the blocks was counterbalanced, i.e. half the participants performed in silence and then with music, and half vice versa. For each block, participants were asked to perform all rumble gestures, prior to performing all

of the everyday activity tasks in a fixed order: open/close mouth, vocalization, drink and swallow, chew gum, and finally do nothing. We counterbalanced the order in which the rumble activities were presented to participants using a balanced Latin square design. For each activity, participants had a training phase prior to the data recording, in which each activity was repeated five times. The study lasted approximately 60 minutes, and participants received a bag of candy as a reward for their participation.

5.3 Data Labelling

We labelled the collected data by hand to be used for analysis and development of the EarRumble recognition pipeline. One researcher precisely identified the start and end time of single and double rumbles, and a second researcher verified the labels. The start time was labelled for the hold rumbles, however the exact end time could not be identified from the recorded data for the majority of samples because we observed that the pressure equalised over time and there were no visible features to identify the end. In total there were 79 hold rumbles that were labelled by hand (49.3%). For repetitive and double rumbling, we labelled the peaks of each rumble to extract the periodicity. The peak detection was automated, using the Python Scipy signal processing library, and hyperparameters were fine tuned if they did not fit the individual samples. The accuracy of the peak detection was verified by visual inspection for all samples.

5.4 Results

The following section explores the participants' perception of ear rumbling and the characteristics of how users contract the tensor tympani muscle when performing different gestures. Figure 5 shows samples for all the different activities that participants performed during the study. All statistical results are reported as significant if $p < .05$, unless using Bonferroni correction to account for multiple comparisons.

5.4.1 Questionnaire Responses. We begin by investigating the responses to the questionnaire that was also administered in section 4. All participants correctly reported that they could rumble, which was validated visually with an otoscope. Five participants self-reported that they heard crackling in addition to the ear rumbling. Fourteen people reported that they could rumble in both ears, with the remaining two reporting that they could only rumble in their right ear. Three people out of the fourteen reported that they could perform ear rumbling in isolation in one ear in addition to both ears (1 right, 2 left). No participants reported that they could perform ear rumbling in both ears and in isolation in both left and right ears. Visual inspection of the pressure sensor data confirmed the laterality that participants reported, with little to no pressure changes being observed when participants reported rumbling in only one ear. All participants reported that they could change the duration of the ear rumble, and that they could repeat the rumble in quick succession. Similarly, all participants reported that they could contract the tensor tympani muscle without having to perform other activities, such as closing their eyes or opening the mouth.

5.4.2 User Perception. Figure 6 shows the median and inter-quartile ranges for the Likert items participants completed during the data

collection (please see Figure 4 for the Likert responses for the lab participants' responses to the questionnaire from section 4). Wilcoxon signed-rank tests showed that there was no significant differences for responses between the music or silence conditions across all the different rumble gestures. In general, participants reported that the ear rumbling gestures were easy to perform on demand, and comfortable. We performed a Friedman test on the "easy to perform on demand" and "comfortable" Likert items to see whether participants' perception was consistent across the different types of ear rumbling gestures. Participant results were significantly different for both the easy to perform ($\chi^2(3) = 11.79, p = .008$), and comfortable Likert items ($\chi^2(3) = 10.22, p = .017$). Pairwise comparisons using the Wilcoxon signed-rank test were performed with a Bonferroni correction for multiple comparisons. These revealed that participants found it significantly easier to perform a single rumble compared with holding the rumble for approximately 1 second ($Z = 4.0, p = 0.0079$). No other results were significant after Bonferroni correction.

5.4.3 Analysis of Ear Rumble Gestures. Figure 7(a)-(c) shows the time to start the ear rumble when performing each of the three different ear rumble gestures (rumble, double rumble, and hold rumble). The start time incorporates the participants' reaction time to the visual stimulus, and the time taken to start the gesture in response. As participants were presented with a five second countdown, we would expect the reaction time to be smaller than that of a random stimulus. We performed statistical tests to see if there was a significant difference between the reaction times across the gestures, however due to the difficulty in extracting ground truth labels for the hold rumble we only compare single versus double rumbles. Shapiro-wilks test of normality revealed both distributions were not normally distributed ($p < .05$), and hence we use the Wilcoxon signed-rank test. There was no significant difference between the start time for the single rumble (Median = 308 ms) compared with the double rumble (Median = 308 ms).

A Wilcoxon signed-rank test ($Z = 407.5, p < .001$) showed that a single rumble (Median = 958 ms) was significantly quicker to perform than a double rumble (Median = 1301 ms) - as to be expected. Interestingly, when analysing the duration of each rumble gesture, we observed how the double rumble (Median: 1010 ms) takes only 48% longer than a single rumble (Median: 684 ms). Figure 7(d) and (e) shows a histogram of the durations for the first and second rumble in the double rumble gesture. By comparing the time difference between rumbles, we see how the first rumble when performed in the double rumble gesture is performed significantly quicker (Median = 419 ms) than the second rumble (Median = 516 ms) using a Wilcoxon signed-rank test ($Z = 425.0, p < .001$) as both distributions were not normally distributed according to the Shapiro-Wilks test of normality. Furthermore, we found that the first ($Z = 313.0, p < .001$) and second rumbles ($Z = 2237.0, p < .001$) were performed significantly more quickly compared with the duration of the single rumble gesture.

Figure 7(e) shows the average time between rumbles for the consecutive rumbling condition. In contrast to the double rumble, we observe that the time difference between rumbles when participants were asked to perform for 5 seconds was significantly longer (Median = 662 ms), and more comparable to the duration of

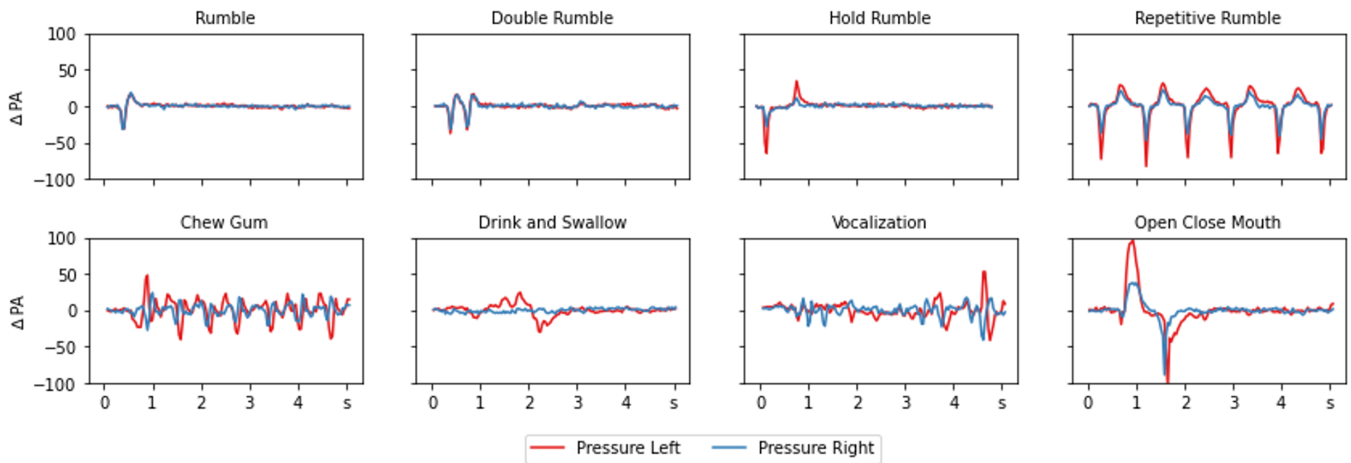


Figure 5: Zero-shifted pressure readings measured in both ears for the different activities. The four rumbling variations show how the pressure readings spike downwards with initial muscle contraction and indicate another peak in the opposite direction after relaxation – creating distinct patterns for rumbling variations. The four noise activities have different characteristics.

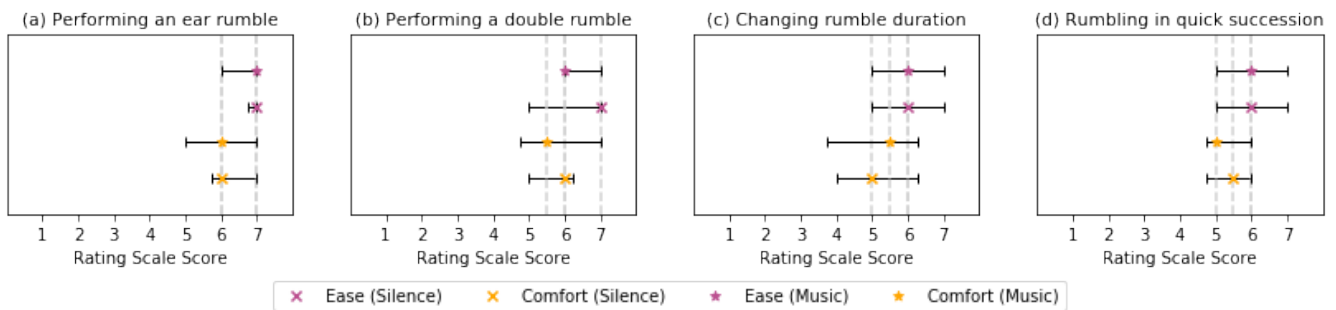


Figure 6: Likert item responses from the 16 lab participants after performing rumble gestures in silence and with music.

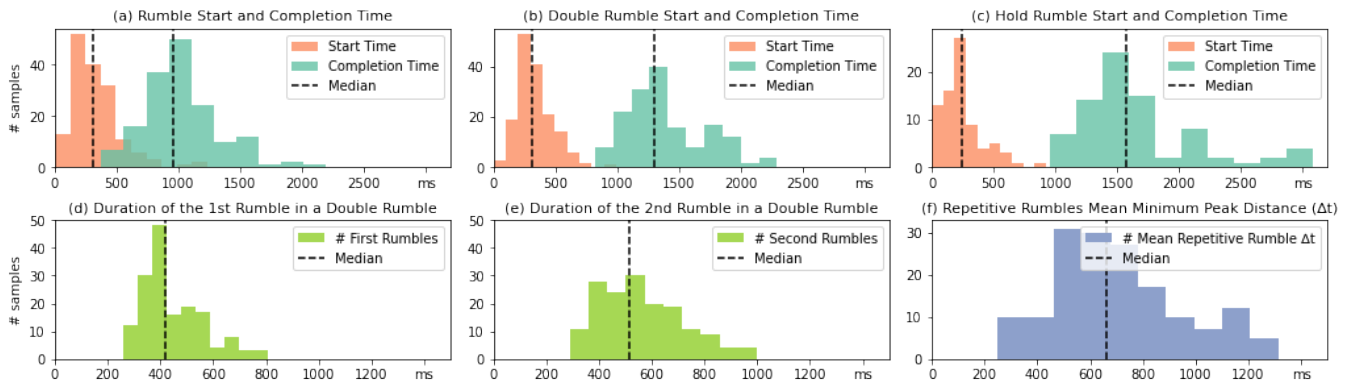


Figure 7: (a)-(c) show the distribution of start and completion time for the different rumbling variations. Interestingly, a double rumble does not take twice as long as a single rumble whereas figure (d) and (e) show how the first rumble is faster to execute than the second. (f) shows how for repetitive rumbling the mean cycle time was significantly longer than during double rumbling.

a single rumble. A Wilcoxon signed-rank test showed there was no statistically significant difference between the single rumble and those performed during the repetition task.

5.4.4 Classifier Performance Evaluation. To assess the performance of different classifiers under ideal conditions, we used the labeled start and end times of the different rumbles and time-constrained noise activities. For the continuous activities we randomly sub-sampled from the recording (1 - 3 seconds). To avoid over-fitting the noise class, we randomly selected evenly per class from all activities. For each classifier candidate, we performed a 5-fold nested cross-validation with grid search for hyper-parameter optimization. Both silent and music conditions were included in the training and test sets.

The best classifier was the XGBoost model which achieved 95% overall accuracy. The optimal hyper-parameters to achieve this were learning rate (0.08), maximum depth (3), and number of estimators (640). Table 1 shows the performance metrics of the different classifiers. The random forest classifier achieves similarly good performance overall, but we found that it did not perform as well on the single rumble class.

Figure 8 shows the confusion matrix of the best classifier. Prolonged and double rumble achieve the best results. The main reason for the confusion between rumbles and noise classes is that rumbles are much less significant in their structure and therefore might be confused easier with short noise samples. Likely, with additional sensors (e.g. IMU or microphone) the noise classes could be discriminated with higher accuracy from the rumble class and vice-versa.

5.4.5 Leave-One Subject Out Validation. There may be variations in the data for the different rumble gestures between participants. The variability could either be temporal (e.g. completion time of rumbles) and/or due to differences in the intensity of the ear rumble (i.e. peak amplitude of the rumbles). Therefore, we trained the best classifier in a leave-one-subject-out cross-validation setting. Overall the classifier achieved 93% overall accuracy (2% decrease), which suggests that the proposed selection of features generalizes well across participants.

5.5 Discussion

We have demonstrated how in-ear barometry can be used to detect contraction of the tensor tympani. The hold rumble gesture was particularly difficult to identify ground truth labels for, and this inherently has ramifications for the development of a classification pipeline because there are fewer samples for training and validation for the hold rumble. It also implies that a large percentage of hold rumbles may not be accurately detected using in-ear barometry. However, participants found contraction of the tensor tympani easy to perform on-demand and comfortable across the different ear rumble gestures.

Ear rumbling can be detected in silence or with music playing, however one participant mentioned that it was harder to focus on the execution of rumbles because the music made the rumbling sound harder to hear which affected the feedback loop. This is something we potentially observed in the previous online questionnaire with the deaf participant and the participant with mild tinnitus, as feedback of the gesture is an important part of the interaction to

notify that their interaction has been successfully registered [8]. However, it is also important to note that in this context there was no action associated with the ear rumbling gestures, and feedback can be provided either indirectly through the response of the system (e.g. changing the song or answering the phone), or directly in response to detection of the ear rumbling itself (e.g. play a sound to indicate rumbling).

There are a number of interesting insights to be gained from the analysis of rumbling characteristics in subsection 5.4.3. The acquisition time of an input technique refers to the time required to acquire the input device so that it is ready for use (e.g. unsheathing a pen), and the homing time refers to the time required to return to a "home" position (e.g. making contact with a finger for touch screen interaction) [23]. Ear rumbling through contraction of the tensor tympani does not introduce any acquisition time, nor any homing time. During the data collection participants were presented with a countdown timer prior to performing the ear rumble gestures, therefore we can see how quickly users are able to respond to the predicted stimulus.

We observed no significant difference for the start times between performing a single or double rumble, however we note that it takes approximately 308 ms to begin the ear rumble. This may be due to the fact that voluntary contraction of the tensor tympani is rarely performed as it serves little purpose. Comparing the time to start a rumble gesture, we note how it is nearly 100 ms shorter than to home in on a device (400 ms) according to the keystroke-level model (KLM) [10]. Based on the duration of a single ear rumble we note how it is similar to typing random letters (500 ms) or complex codes (750 ms) according to KLM.

Our insights into repetitive rumbling indicate that contraction of the tensor tympani could be suitable for more complex interactions than the three basic gestures we have explored here, and opens up the opportunity to use ear rumbling with switch scanning interfaces [49], rhythmic patterns [20], or beat synchronisation [53, 57]. Our results provide insights into what kind of tempo one could use for these interactions to optimise throughput of the technique, and participants reported that this was generally easy to perform on-demand and comfortable. There is also scope to incorporate ear rumbling interaction with existing techniques, such as extending Ando et al.'s in-ear barometry-based gesture set [3], so that a wider vocabulary is available for users when interactions beyond simple binary choices are required.

6 USABILITY EVALUATION

We performed a study to explore the performance and usability of EarRumble as a hands- and eyes-free input technique. Using a real-time implementation of the pipeline featured in subsection 3.3, we ground our exploration in three manual, dual task application scenarios that one might face when using earables in everyday life. The goals of the study were to test how well the pipeline worked, and to gather feedback from users when using EarRumble for interaction.

6.1 Participants and Apparatus

We invited eight participants (7 male, 1 female, Age: $M = 24.5$ $SD = 2.7$) from the previous data collection study in section 5. Participants

Classifier	Pre.	Rec.	F1	Acc.
Dummy	0.28	0.28	0.28	0.28
RBF Kernel SVM	0.81	0.79	0.80	0.79
kNN	0.82	0.81	0.81	0.81
Random Forest	0.95	0.95	0.95	0.95
XGBoost	0.95	0.95	0.95	0.95

Table 1: Performance metrics comparison of different classifiers. XGBoost yielded the best overall performance for the individual rumbling activities.

were seated in a regular office chair in front of a desktop PC. The pressure sensors in the earphones were connected via USB to a desktop computer which was running the EarRumble detection pipeline software in Python. Participants used the custom-built EarRumble earphones described in section 3, and a separate pair of HolyHigh in-ear earables which features a mechanical click button on the outside of both earphones. Participants were instructed to use either earphone for the click, depending on their handedness preference.

6.2 Design

We compared three rumble gestures (single rumble, double rumble, and hold rumble) using the EarRumble technique with a simple button click on a pair of smart earphones using three analogous gestures (click, double click, hold). A button click was chosen instead of a tap gesture because the button allows for the accurate detection of hold gestures. We chose three different use cases using two applications for the evaluation. These were chosen to evaluate the techniques in the context in which they may be used in real-life.

6.2.1 Incoming Call. The first application scenario featured an incoming call, whereby a ringtone plays in the earphones and the user can either accept (rumble/click), reject (double rumble/click), or mute (hold rumble/button) the incoming phone call. We mimicked a use case in which the user’s attention is on a manual task. We gave participants a primary task of typing a piece of text on the desktop PC whilst playing music in the background through the earphones. The phone would then ring and the user would be tasked with either accepting, rejecting, or muting the phone call, before carrying on with the typing task.

6.2.2 Audio Player. The second and third application scenarios feature an audio player in which the user can play and pause (rumble/click), skip to the next track (double rumble/click), or go back to the previous track (hold rumble/button). The second scenario consists of an audio transcription task whereby participants are tasked with transcribing sentences being read aloud through the earphones. After each sentence the participant is required to pause the audio, write the sentence down, and resume playback to hear the next sentence. The third scenario consisted of a music playlist in which participants were required to skip forwards or backwards in order to find specific songs.

True Label	Predicted Label			
	Rumble	Double Rumble	Prolonged Rumble	Noise
Rumble	0.86	0.04	0.03	0.07
Double Rumble	0.01	0.97	0.01	0.01
Prolonged Rumble	0.00	0.01	0.98	0.01
Noise	0.05	0.01	0.01	0.93

Figure 8: The confusion matrix shows how short, sub-sampled noise activities can confuse the classifier for single rumble detection and vice versa.

6.3 Procedure

Participants began the study by completing demographic information and signing a consent form. They were then given one of the application tasks to practice with, using both input techniques. The order in which the applications were presented to participants were counterbalanced, as was the order of the input techniques.

For the incoming call scenario, participants completed the task in three blocks. For each block, the participant had to accept/reject/mute all incoming calls. Only one action was chosen per block to reduce the burden of memorising which action to take. In each block, the participant received four phone calls at intervals of 20 seconds. This was chosen to give the participant time to resume typing, and long enough to reduce the chance of precisely predicting when a call would occur. For the audio player scenario, participants were given four sentences to transcribe, resulting in eight rumbles/clicks in order to pause and resume playback. The song play list consisted of five songs, and participants were tasked with finding the songs which involved skipping 3x forward, 2x back, 1x forward, and 2x back – resulting in four gestures each of the double rumble/click and hold rumble/click.

For the incoming call task we measured the response time of participants from when the call was triggered to the corresponding action. We also measured the time taken to return to the typing task, defined as the first keystroke after the incoming call had been actioned (e.g. rejected). Participants were asked to report any errors during the interactions arising from (a) incorrect detection of gestures (e.g. rumbles/clicks which aren’t detected or which should have been), (b) incorrect classification of gestures (e.g. detection of a single rumble/click instead of a hold rumble/click), and (c) user error (i.e. performing the wrong gesture). Participants completed the NASA TLX [21] and were asked what they liked and disliked about each input technique for both of the applications. After both techniques had been performed for an application, participants were asked which they preferred using for the specific application, and at the end of the study they were asked which one they preferred overall.

6.4 Results

When asked about their overall preferences, six participants preferred the EarRumble technique, and two preferred the button click. The EarRumble technique was favoured because it required less effort (P2, P4, P6, P8), and was more comfortable (P5). P3 specifically highlighted the single ear rumble gesture was their preferred technique. The button click was preferred because it was more robust

(P1), and P7 noted that they would have preferred the earphones if it was tap input rather than button click because the EarRumble was lacking immediate feedback. We discovered issues with reliably detecting the hold rumble – only 56% of prolonged rumbles were detected correctly compared with 91% of single rumbles and 94% of double rumbles throughout all tasks and across participants. We ran a Friedman tests on the responses to the NASA TLX but found no significant differences between the input technique and task conditions. In the following we discuss the feedback from the two application scenarios.

6.4.1 Phone Call Task. For the phone call task, seven participants highlighted the advantage of not having to take their hands off the keyboard, and ability to continue typing immediately. P6 perceived that it felt much quicker than the button to interact. Excluding erroneous detections, the mean time between the phone call and detection of gesture was 3.40 s for EarRumble, and 3.34 s for the button click. The mean time between detection and the first key-stroke during the call task was 1.09 s for EarRumble and 1.51 s for the button click. We did not run statistical tests due to the different numbers of successful detections and low participant numbers. Interestingly, P1 also noted how it was nice not to have to use voice – alluding to the social acceptability of the technique. However, six participants struggled using the hold rumble for interaction. P1, P5, and P7 reported that the hold rumble required additional concentration and was more uncomfortable to perform. P4 also commented on the latency of the rumble detection, and both P1 and P7 noted the lack of immediate feedback from the EarRumble techniques.

The participants also saw advantages of using the button for input, because it was clear how it worked and have used it before (P1, P7), provided immediate haptic feedback (P1, P4, P8), and worked reliably (P1, P3) with low latency (P1). However, 3 participants highlighted a disadvantage of having to take their hands off the keyboard to interact (P1, P2, P4) which they felt broke their workflow. Also, five participants felt the physical click button was hard to press and hurt the ear canal (P2, P3, P5, P6, P7). Two preferred the button click because of the immediate feedback (P1) and because it was more robust (P7).

Six out of the eight participants preferred the EarRumble technique for the phone call task, despite the technical issues with the hold rumble gesture. Participants preferred EarRumble because it required no extra movement (P2, P6), less time (P5), and allowed them to continue typing (P3, P8). P4 preferred the technique because “it felt magical” and “almost felt telepathic”.

6.4.2 Audio Player Tasks. Participants gave similar feedback for the audio player tasks. The EarRumble technique was perceived to be faster because it does not require the use of the hands (P1, P3, P4, P8), and was low effort (P5). P6 described the technique as “much more practical” than the button click. Interestingly, two participants described the interactions with the audio player using EarRumble as “fun” (P7, P8). The disadvantages cited once again referred to the hold rumble (P1, P2, P5, P6, P7, P8), and higher latency of rumbling detection (P3, P4, P5).

Feedback was similar as well for the button click with the audio player, with it being described as a known technique (P5), very robust (P1), and immediate haptic feedback was an advantage (P1,

P4, P7, P8). Participants described it as “annoying” to take the hands off the keyboard when pausing the text to write (P1, P3, P4), and once again they highlighted that the click hurts the ear canal (P2, P3, P5, P6, P7, P8).

No participants changed their preference between phone call task and audio player task. Participants preferred the EarRumble technique because it was low effort (P3, P4, P8), and because it does not require the use of the hands (P6). P2 said they preferred EarRumble because they think it would be “perfect when listening to music in bed”. P5 reported that they preferred EarRumble because the button click was uncomfortable to use, and the remaining participants preferred the button because it was more robust (P1, P7).

6.5 Discussion

This usability evaluation highlights the low-effort, hands-free nature of the EarRumble technique, which was the main motivation of adopting ear rumbling for interaction. Scenarios in which participants could imagine using the EarRumble technique included during focused work (P1, P2, P4), when hands are occupied (P2, P5, P8), for secretive input (P2, P3), to interact without any noise, e.g. speech (P3), or for music or calls (P6). However, as expected from the results in section 5, the pressure sensing technique did not reliably detect the hold rumble gestures. For some users, the accuracy was very high (90+%), however there are larger issues around detecting the hold rumble using in-ear barometry. We also note how the latency of the pipeline was an issue for some participants, which with further optimisation could be further reduced allowing for quicker selection times.

7 FUTURE WORK AND LIMITATIONS

EarRumble requires air-tight sealing of the user’s ear canal. Blocking the ear canal with headphones for prolonged time can change the “climate” of the ear (e.g. temperature and humidity) and is often said to support the entry of, e.g., bacteria in the middle ear. However, no significant clinical evidence exists to back increased bacterial or fungal exposure by continuous use of regular earphones [13, 37]. Regarding comfort, we had one participant mention that they felt uncomfortable sealing the ear canal for prolonged periods, and this was also applicable when wearing regular in-ear headphones (section 6-P7). The ear caps of the current EarRumble system are foam soft type plugs which fit in tightly in the ear canal and go deeper into the ear canal than regular ear plugs, meaning they may not feel as comfortable after expansion as regular in-ear type plastic caps. Initially, we experimented with standard headphone plastic caps and could observe pressure changes clearly and consistently, however in the initial exploration of in-ear barometry we noticed that some users could not perform the ear rumbles as strongly and they were harder to detect. Nevertheless, standard in-ear caps should be investigated further in future work for real-world applications.

Ear rumbling is a prime candidate for providing simple gestures on the go – for example, interaction in a crowded train would be easily possible without requiring any movement by the user, or other cases where mobility is limited. We only investigated the use of ear rumbling with users sat down in a stationary position, and did not investigate the social acceptability implications of the technique.

Real-world deployments of the technology may reveal interesting insights into how movement in unconstrained environments affects the sealing of the ear canal, the detection pipeline, and/or a user's ability to contract the tensor tympani. It also may reveal how many other actions throughout the day could lead to false positives (e.g. yawning), as the false positive actions we chose in this paper were based on easily replicable actions that most closely resembled the contraction of the tensor tympani. There may also be scope to suppress false positives through the use of other sensing modalities, e.g. sensing chewing gum using an IMU [33].

The applied sensing principle allows for reliable detection of single and double rumbles, however the hold rumbles proved to be problematic and the current setup does not allow to derive the duration of rumbles precisely. In the future, other sensing principles may be used to realize ear rumble detection more reliably, e.g. by an in-ear camera or acoustic impedance measurements [40, 41]. The latter might be even realized with off-the-shelf hardware with noise canceling earphones as they have an in-ear microphone to for the noise canceling feedback loop [1].

Finally, we could not identify any literature discussing the consequences of voluntary tensor tympani contraction over a long-term basis. We discussed potential long-term safety concerns with the ear, nose, and throat doctor consulted during the project, who noted no known safety issues with voluntarily contracting the tensor tympani and could not see why this would cause any problems. However, the absence of data relating to this does not imply long-term safety, which future work should further investigate.

8 CONCLUSION

EarRumble uses in-ear barometry to detect the contraction of the tensor tympani muscle, known as ear rumbling, allowing users to provide low-effort, discreet interaction using earable devices. An online questionnaire showed that 44% of respondents reported that they could perform ear rumbling, and a data collection with 16 participants provided insights into the level of control users have over contracting the tensor tympani, demonstrating how ear rumbling is a viable interaction technique. We explored how interaction could be achieved using three simple “gestures” using a detection pipeline consisting of feature extraction and gradient boosted classification. Single and double rumbles could be accurately detected, however detection of a rumble that is held for a prolonged period (e.g. 1 second) proved to be problematic for many participants due to the pressure sensing approach. A usability evaluation grounded in three manual, dual task application scenarios showed the low-effort, hands-free advantages of the technique relative to providing input via a button on the earable. The use of ear rumbling has the potential to be useful in a number of application scenarios involving on-the-go mobile interaction, with scope for future work to investigate more robust sensing techniques.

REFERENCES

- [1] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial expression recognition using ear canal transfer function. In *Proceedings of the 23rd International Symposium on Wearable Computers*. 1–9.
- [2] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of chewing sounds for dietary monitoring. In *International Conference on Ubiquitous Computing*. Springer, 56–72.
- [3] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 679–689.
- [4] RD Angeli, M Lise, CC Tabajara, and TB Maffacoli. 2013. Voluntary contraction of the tensor tympani muscle and its audiometric effects. *The Journal of Laryngology and Otology* 127, 12 (2013), 1235.
- [5] Takayuki Arakawa, Takafumi Koshinaka, Shohei Yano, Hideki Irisawa, Ryoji Miyahara, and Hitoshi Imaoka. 2016. Fast and accurate personal authentication using ear acoustics. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 1–4.
- [6] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jimenez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An exploration of tooth click gestures for hands-free user interface control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 158–169.
- [7] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015. Stick it in your ear: Building an in-ear jaw movement sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 1333–1338.
- [8] Victoria Bellotti, Maribeth Back, W Keith Edwards, Rebecca E Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 415–422.
- [9] Vincent Buil, Gerard Hollemans, and Sander van de Wijdeven. 2005. Headphones with touch control. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. 377–378.
- [10] Stuart K Card, Thomas P Moran, and Allen Newell. 1980. The keystroke-level model for user performance time with interactive systems. *Commun. ACM* 23, 7 (1980), 396–410.
- [11] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y. Chen, Wen-Huang Cheng, and Bing-Yu Chen. 2013. FingerPad: Private and Subtle Interaction Using Fingertips. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) (*UIST '13*). Association for Computing Machinery, New York, NY, USA, 255–260. <https://doi.org/10.1145/2501988.2502016>
- [12] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package). *Neurocomputing* 307 (2018), 72–77.
- [13] SJ Cooper. 1985. Relationship of hearing protector type and prevalence of external auditory canal pathology. In *Am. Ind. Hyg. Asso. Conf. Las Vegas, NV, paper*, Vol. 23.
- [14] Enrico Costanza, Samuel A Inverso, and Rebecca Allen. 2005. Toward subtle intimate interfaces for mobile devices using an EMG controller. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 481–489.
- [15] Alexander De Luca, Roman Weiss, and Heiko Drewes. 2007. Evaluation of Eye-Gaze Interaction Methods for Security Enhanced PIN-Entry. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces* (Adelaide, Australia) (*OZCHI '07*). Association for Computing Machinery, New York, NY, USA, 199–202. <https://doi.org/10.1145/1324892.1324932>
- [16] Artem Dementyev and Joseph A Paradiso. 2014. WristFlex: low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 161–166.
- [17] Gisle Djupesland. 1964. Middle ear muscle reflexes elicited by acoustic and nonacoustic stimulation. *Acta oto-laryngologica* 57, sup188 (1964), 287–292.
- [18] Euan Freeman, Gareth Griffiths, and Stephen A. Brewster. 2017. Rhythmic Micro-Gestures: Discreet Interaction on-the-Go. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) (*ICMI '17*). Association for Computing Machinery, New York, NY, USA, 115–119. <https://doi.org/10.1145/3136755.3136815>
- [19] Pablo Gallego Cascón, Denys J.C. Matthies, Sachith Muthukumarana, and Suranga Nanayakkara. 2019. ChewIt. An Intraoral Interface for Discreet Interactions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300556>
- [20] Emilien Ghomi, Guillaume Faure, Stéphane Huot, Olivier Chapuis, and Michel Beaudouin-Lafon. 2012. Using rhythmic patterns as an input method. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1253–1262.
- [21] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [22] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1063–1072.
- [23] Ken Hinckley and Daniel Wigdor. 2002. Input technologies and techniques. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (2002), 151–168.

- [24] Hans-Eric Holst, Sven Ingelstedt, and Urban Örtengren. 1963. Ear drum movements following stimulation of the middle ear muscles. *Acta Oto-Laryngologica* 56, sup182 (1963), 73–83.
- [25] Lei Jing, Zixue Cheng, Yinghui Zhou, Junbo Wang, and Tongjun Huang. 2013. Magic Ring: a self-contained gesture input device on finger. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*. 1–4.
- [26] Heath G Jones, Nathaniel T Greene, and William A Ahroon. 2019. Human middle-ear muscles rarely contract in anticipation of acoustic impulses: Implications for hearing risk assessments. *Hearing research* 378 (2019), 53–62.
- [27] Christina Katsini, Yasmeen Abdrabou, George E Raptis, Mohamed Khamis, and Florian Alt. 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [28] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H Thomas. 2017. EarTouch: turning the ear into an input surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–6.
- [29] Ingmar Klockhoff. 1961. Middle ear muscle reflexes in man. *Journal of Occupational and Environmental Medicine* 3, 10 (1961), 486.
- [30] Juyoung Lee, Hui-Shyong Yeo, Murtaza Dhuliawala, Jedidiah Akano, Junichi Shimizu, Thad Starner, Aaron Quigley, Woontack Woo, and Kai Kunze. 2017. Itchy Nose: Discreet Gesture Interaction Using EOG Sensors in Smart Eyewear. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (Maui, Hawaii) (ISWC '17)*. Association for Computing Machinery, New York, NY, USA, 94–97. <https://doi.org/10.1145/3123021.3123060>
- [31] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. 2013. Earput: Augmenting behind-the-ear devices for ear-based interaction. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1323–1328.
- [32] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-Air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (Charlotte, NC, USA) (UIST '15)*. Association for Computing Machinery, New York, NY, USA, 63–71. <https://doi.org/10.1145/2807442.2807489>
- [33] Roya Lotfi, George Tzanetakis, Rasit Eskicioglu, and Pourang Irani. 2020. A comparison between audio and IMU data to detect chewing events based on an earable device. In *Proceedings of the 11th Augmented Human International Conference*. 1–8.
- [34] Hiroyuki Manabe and Masaaki Fukumoto. 2011. Tap control for headphones without sensors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 309–314.
- [35] Denys JC Matthies. 2013. InEar BioFeedController: a headset for hands-free and eyes-free interaction with mobile devices. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1293–1298.
- [36] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. 2017. Earfieldsensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1911–1922.
- [37] R Mazlan, L Saim, A Thomas, R Said, and B Liyab. 2002. Ear infection and hearing loss amongst headphone users. *The Malaysian journal of medical sciences: MJMS* 9, 2 (2002), 17.
- [38] Edward D McCoul, Vijay K Anand, and Paul J Christos. 2012. Validating the clinical assessment of eustachian tube dysfunction: The eustachian tube dysfunction questionnaire (ETDQ-7). *The Laryngoscope* 122, 5 (2012), 1137–1141.
- [39] Christian Metzger, Matt Anderson, and Thad Starner. 2004. Freedigiter: A contact-free device for gesture control. In *Eighth International Symposium on Wearable Computers*, Vol. 1. IEEE, 18–21.
- [40] Aage R Møller. 1961. Network model of the middle ear. *The Journal of the Acoustical Society of America* 33, 2 (1961), 168–176.
- [41] Aage R Møller. 1974. The acoustic middle ear muscle reflex. In *Auditory system*. Springer, 519–548.
- [42] Henning Pohl, Andreea Muresan, and Kasper Hornbæk. 2019. Charting Subtle Interaction in the HCI Literature. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3290605.3300648>
- [43] Adam Politzer. 1902. *A Text-Book of the Diseases of the Ear for Students and Practitioners*. Baillière, Tindall and Cox.
- [44] Adam Politzer. 1909. *A Text-book of the Diseases of the Ear: For Students and Practitioners*. Lea Bros.
- [45] Mark D. Reilly, Haifeng Shen, Paul R. Calder, and Henry Been-Lirn Duh. 2013. Understanding the Effects of Discreet Real-Time Social Interaction on Student Engagement in Lectures. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (Adelaide, Australia) (OzCHI '13)*. Association for Computing Machinery, New York, NY, USA, 193–196. <https://doi.org/10.1145/2541016.2541086>
- [46] Julie Rico and Stephen Brewster. 2009. Gestures all around us: user differences in social acceptability perceptions of gesture based interfaces. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–2.
- [47] Himanshu Sahni, Abdelkareem Bedri, Gabriel Reyes, Pavleen Thukral, Zehua Guo, Thad Starner, and Maysam Ghovanloo. 2014. The tongue and ear interface: a wearable system for silent speech recognition. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*. 47–54.
- [48] Gerhard Salomon and Arnold Starr. 1963. Electromyography of middle ear muscles in man during motor activities. *Acta Neurologica Scandinavica* 39, 2 (1963), 161–168.
- [49] Richard C Simpson and Heidi Horstmann Koester. 1999. Adaptive one-switch row-column scanning. *IEEE Transactions on Rehabilitation Engineering* 7, 4 (1999), 464–473.
- [50] Susan Standring. 2015. *Gray's anatomy e-book: the anatomical basis of clinical practice*. Elsevier Health Sciences.
- [51] Ravi Vaidyanathan, Beomsu Chung, Lalit Gupta, Hyunseok Kook, Srinivas Kota, and James D West. 2007. Tongue-movement communication and control concept for hands-free human-machine interfaces. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37, 4 (2007), 533–546.
- [52] Tomás Vega Gálvez, Shardul Sapkota, Alexandru Dancu, and Pattie Maes. 2019. Byte.It: Discreet Teeth Gestures for Mobile Device Interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3290607.3312925>
- [53] Eduardo Velloso, Marcus Carter, Joshua Newn, Augusto Esteves, Christopher Clarke, and Hans Gellersen. 2017. Motion correlation: Selecting objects by matching their movement. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 3 (2017), 1–35.
- [54] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, and Yuanchun Shi. 2019. EarTouch: facilitating smartphone use for visually impaired people in mobile and public scenarios. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [55] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-dependent requirements. In *IFIP conference on human-computer interaction*. Springer, 559–575.
- [56] Xuhai Xu, Haitian Shi, Xin Yi, Wenjia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [57] Tengxiang Zhang, Xin Yi, Ruolin Wang, Yuntao Wang, Chun Yu, Yiqin Lu, and Yuanchun Shi. 2018. Tap-to-Pair: Associating Wireless Devices with Synchronous Tapping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–21.