A Taxonomy of Noise in Voice Self-reports while Running

Tao Bi
University College London
London, United Kingdom
t.bi@ucl.ac.uk

Temitayo Olugbade
University College London
London, United Kingdom
temitayo.olugbade.13@ucl.ac.uk

Akhil Mathur
Nokia Bell Labs
Cambridge, United Kingdom
akhil.mathur@nokia-bell-labs.com

Catherine Holloway
University College London
London, United Kingdom
c.holloway@ucl.ac.uk

Aneesha Singh
University College London
London, United Kingdom
aneesha.singh@ucl.ac.uk

Enrico Costanza
University College London
London, United Kingdom
e.costanza@ucl.ac.uk

Nadia Berthouze
University College London
London, United Kingdom
nadia.berthouze@ucl.ac.uk

ABSTRACT

Smart earables offer great opportunities for conducting ubiquitous computing research. This paper shares its reflection on collecting self-reports from runners using the microphone on the smart eSense earbud device. Despite the advantages of the eSense in allowing researchers to collect continuous voice self-reports anytime anywhere, it also captured noise signals from various sources and created challenges in data processing and analysis. The paper presents an initial taxonomy of noise in runners' voice self-reports data via eSense. This is based on a qualitative analysis of voice recordings based on eSense's microphone with 11 runners across 14 in-the-wild running sessions. The paper discusses the details and characteristics of the observed noise, the challenges in achieving good-quality self-reports, and opportunities for extracting useful contextual information. The paper further suggests a noise-categorization API for the eSense or other similar platforms, not only for the purpose of noise-cancellation but also incorporating the mining of contextual information.

UbiComp/ISWC '22 Adjunct, September 11–15, 2022, Cambridge, United Kingdom
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9423-9/22/09…$15.00
https://doi.org/10.1145/3544793.3563421

CCS CONCEPTS

• Human-centered computing • Human computer interaction (HCI) • Empirical studies in HCI

KEYWORDS

Taxonomy; Voice; Self-reports; Runners; Running; In-the-wild Experience Sampling; Smart earbuds; Earables; eSense;

ACM Reference format:

Tao Bi, Temitayo Olugbade, Akhil Mathur, Catherine Holloway, Aneesha Singh, Enrico Costanza, Nadia Berthouze. 2022. A Taxonomy of Noise in Voice Self-reports while Running. *In Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct), September 11--15, 2022, Cambridge, United Kingdom, 5 pages.* https://doi.org/10.1145/3544793.3563421

1 INTRODUCTION

Smart earables offer great opportunities for experience sampling and collection of user self-reports [1] in ubiquitous contexts. Traditional experience sampling methods (ESM) that use mobile phone based applications require people to physically interact with a phone multiple times [2]. This is not very practical when collecting self-reports from runners while running because runners use their hands and arms to maintain balance and movement flow. Smart earbuds can offer a hands-free experience for people to self-report via voice recordings [3]. Runners receive ESM prompts via earbud speakers and use speech to self-report back to the earbud microphone, which creates less physical or biomechanical interference with the runner's body movement [4].

Also, earbuds are lightweight and widely accepted by many runners to consume music during running.

In our study, eSense was used to deliver an ESM schedule and to record verbal self-reports of feelings of runners at run time in the wild. eSense is a multi-sensory earable platform that is widely used in the HCI research community [5, 6]for collecting audio recordings via its embedded microphone. Overall, 11 runners (five males, and six females) had a total of 14 running sessions. Nine of the runners completed only one session, one runner completed two sessions, and another runner completed three sessions. The duration of running sessions (and eSense data recording) averaged 34.7 ± 15.1 minutes. The researcher guided participants during the 5-min trail session. The study did not control the running environment, runner ability, or runner performance. Participants were given with flexibility to run anywhere at any self-selected pace. For more details on the data collections and other aspects of the work see [3, 4].

Despite the earbuds offering a more ubiquitous and less intrusive ESM, we faced challenges in data processing and analysis due to the noise in the captured audio data. To characterize the noise sources and their information value, we used Nvivo [7] to label all potential noise information and then qualitatively analyzed the characteristics of all potential noise information. The noise coding process is mainly based on the first author's listening and interpretation of audio. It also used the author's observation and contextual notes taken during data collection process, i.e.., observation of the runner's apparel and accessories through the study session when participants run indoor or at reachable distance outdoor. The following sections will present an initial taxonomy of the noise observed in the runner voice self-reports.

2 A TAXONOMY OF NOISE IN RUNNERS' VOICE SELF-REPORTS

As shown in Table 1, this taxonomy consists of 14 noise categories: Car noise, Traffic noise, Sound of foot strike, Outdoor terrain noise, Earbuds rubbing noise, Breathing sound, Clothes rubbing noise, Wind noise, Personal item vibration noise, Treadmill machine noise, Foot strike and treadmill impact sound, Animal sound, Passenger talking, and Shop's speaker noise.

Each noise category is associated with its relevant running context (e.g., road run, gym treadmill run). For each category, the taxonomy also includes the characteristics and factors that influence the impact of the noise (e.g., pitch, volume, wind direction). It also classifies whether the noise is synchronized with the run, i.e. whether or not the noise is concomitant with the run. In addition, each category is evaluated with respect to its impact on the quality of voice self-reports. For example, a noise category is labelled as high impact if it severely renders the voice recording difficult to be understood and hence transcribed.

The noise category itself as well as the characteristics of the noise can be useful contextual info for understanding the noise experience, e.g., outdoor surface vs treadmill surface enables detection of running context, animal sounds might explain references to animals in self-report, frequency of foot strike can be useful in capturing fatigue or pacing, etc.

Table 1: Taxonomy of noise in runners' voice self-reports

Noise type	Characteristics & Factors	Synchronized with running?	Running context	Impact on speech recognition
Car noise	Scrapping or chirping sound when a car passes; Increasing pitch when the car is approaching; Decreasing pitch when the car is driving away; Horning sound;	No	Road run	High
Traffic noise	Motorbike engine noise; Car engine noise; Police car or emergency car alarm sound	No	Road run Park run	High
Sound of foot strike	High-rhythm sound during a fast run; Low-rhythm sound during a slow run; High volume sound on heavy landing; Low volume sound on light landing; Heel-to-front transition sound	Yes	Road run Gym run	High
Outdoor terrain noise	Road; Running Track; Grass; Trail; Treadmill	Yes	All run	High
Earbuds rubbing noise	sSense earbuds are unstable in intense physical activities. The rubbing noise has a very high frequency.	Yes	All run	High
Breathing sound	Inhale and exhale sound different.	Yes	All run	High
Clothes rubbing nose	Same frequency & rhythm as food strike. Chuffing noise from arm, hands, legs	Yes	All run	High
Wind noise	Headwind; downwind; crosswind Wind direction causes different noise levels. Wind direction interacts with voice volume.	No	Outdoor run	High
Personal item vibration noise	Keys; Earrings; Necklaces	Yes	All run	High
Treadmill machine noise	Treadmill belt noise: squeaking, screeching and whining noises; Treadmill motor noise; High frequency & High volume; Other gym machine noise	Yes	Gym run	High
Foot strike and treadmill impact sound	Landing sound volume; Soft or heavy landing; High speed creates high pitched noise from treadmill machine;	Yes	Gym run	High
Animal sound	Birds sound, e.g., seagulls	No	Outdoor run	Low
Passenger talking	Group talk vs individual talk lower frequency voice vs higher frequency voice	No	Road run Gym run	High

3 DEMONSTRATIONS OF AUDIO AND NOISE SIGNALS IN A RUNNER'S VOICE SELF-REPORTS

The taxonomy above shows that the noise observed had different characteristics and were typically high impact. Here, we further provide visual representations of the raw audio highlighting the noise signals, using the software Audacity [8], based on the data captured from one of the running sessions introduced in Section 2 as a case study. These visual representations aim to 1) illustrate not only how different noise categories and their characteristics impact the raw audio signals, and 2) illustrate how raw audio and noise signals could infer meaningful information.

Figure 1 shows the full voice recording for the running session. In this recording, all sounds (voice and noise signals) overlapped, which makes it difficult to recognize a runner's speech. Figure 2 shows an excerpt (approximately 3 minutes) of the full voice recording in Figure 1. In Figure 2, we distinguished between noise signals and speech. Voice signals when the runner was talking (segment C in the figure) occur throughout the excerpt. Meanwhile, there are aural occlusions at several points, e.g., from noise of one foot (segments A in the figure) or other (segments B in the figure) landing. The feet strike noise (segments A and B) overlaps with the voice signal (segment C). Further, the noise (segments A and B) are of higher volume than the voice (segment C), which makes it challenging to extract the voice content from the overlapped area.

Beyond the negative implications of the noise, we deduced potentially valuable information from them. For instance, as can be seen in Figure 2, segment A always has higher volume than segment B, which suggests that this runner may have a higher landing impact in one foot than in the other.

4 DISCUSSIONS

As shown in Table 1, noise in runner voice self-report is complex but rich in information. Although we applied simple noise reduction techniques offered by Audacity [9], the results were not satisfactory enough to produce a good-quality voice recording. This is due to the irregular pattern of some of the noise such as wind noise, traffic noise, and noise from road surfaces with variations. In addition, the volumes of step and strike noise were much higher than the runner's voice. The noise from the earbuds' rubbing was even stronger as it was the closest to the microphone. Such factors and characteristics make it challenging to apply existing noise reduction techniques. While our analysis and discussion is limited in that we did not review how existing noise cancellation technologies may address the challenges that the observed noise and their characteristics pose, our findings outline critical issues that arise in running scenarios, some of these issues may be unique to ESM in running and may not be observed in other use cases (e.g. listening to music during outdoor walking,

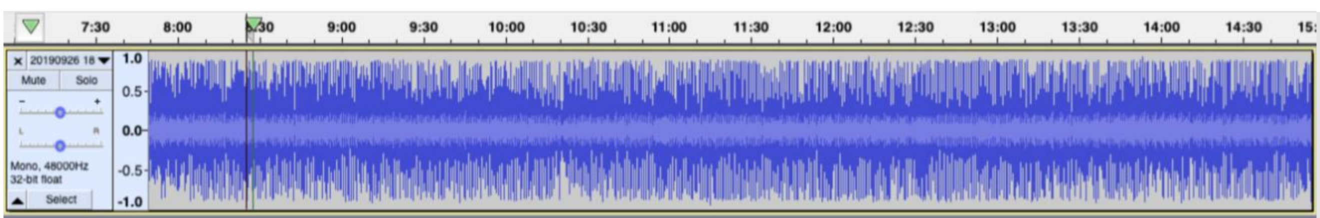

Figure 1: An overview of a runner's voice self-report while running.

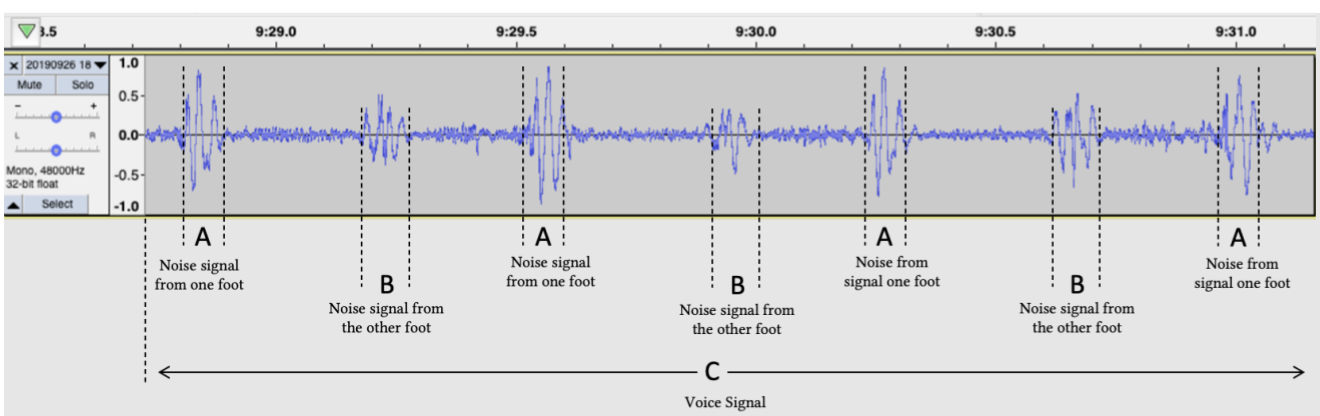

Figure 2: An excerpt of a runner's voice self-report.

private phone call in a public space) that noise cancellation technologies will typically target.

The proposed taxonomy sheds a light on what type of noise existed in the running context, and the summary of features of each noise category can be useful to guide development of noise reduction API (or real-time noise cancellation software) that address the specific scenario of running ESM or ESM in the wild in general. For instance, as highlighted in Table 1, headwind, downwind, or crosswind, not only due to the natural wind direction but also the running direction of a runner themselves, degrades the quality of voice signal as wind vibrates air particles that the voice vibrates [10]. Further, while crosswind direction could additionally alter the direction of voice transmission to one side; headwind could collide with the voice, more than tailwind. Noise cancellation and voice augmentation techniques could better accommodate such factors, to help researchers generate a better-quality voice recording via eSense or other similar platforms.

However, noise can have value, particularly in providing contextual information and personal affective experience. For instance, wind noise can provide information useful for contextualizing the runner's experience self-report. Headwind noise signals can be different from downwind noise, whereas headwind can contribute to making a run extremely difficult. In the transcribed voice self-reports, a runner for example said "fight the headwind" which could indicate that they were struggling or putting much higher effort into the moment. On the contrary, a runner "feels a bit easier now, with the downwind". Here is another example, a runner said, "I am cold, I'm going downhill so the wind catches me, I'm sure it will get better shortly". Such wind direction can be a factor that makes a runner perceive the run as difficult. In such a situation, runners might benefit from some sort of digital cheering that can be delivered via earbuds. On another hand, headwind can also be a feel-good factor in the run. For instance, a runner referred to wind as "nice to run into a breeze on a hot day". When the temperature is high, but wind is gentle and cool, this could be a moment in which a runner mentally enjoys the run. Therefore, wind noise can not only serve as a contextual measurement but also a measurement for personal affective experience. Similarly, car noise can infer traffic and be used as a safety measurement. Digital reminders could be sent to runners to be aware of traffic if they are running on the road. Foot strike noise even has more information that can be used to detect a runner's running performance such as cadence and stride type (heel stride, middle stride, front stride). It could also be used to infer a runner's mental state again. For example, fatigue could make a runner switch to more heel strikes, which is associated with a higher impact on the ground[11].

5 CONCLUSIONS

This paper reports a preliminary taxonomy of noise observed in voice recordings from runners during verbal self-report, based on a qualitative analysis of the audio recordings. The taxonomy includes the noise types, their characteristics and factors, whether the noise type is synchronized with running or not, and how bad the noise affects the speech recognition quality. Our findings highlight need for noise investigation before audio data collection in the wild. Our taxonomy further highlights that noise can be both negative and positive factors. While noise clearly undermines extraction of the desired self-report of experience, it has the potential to provide rich contextual information for deeper understanding of the experience. eSense or other similar smart earbuds platforms will be more valuable when equipped with modules (e.g., API) capable of noise reduction, noise categorization, and extraction of contextual information from non-verbal aural signals. The research in this paper is still in early stages and the aim of this paper is to prompt discussion within/across relevant areas (human-computer interaction, audio engineering, machine learning).

ACKNOWLEDGMENTS

Tao Bi was supported by a studentship deriving from a grant awarded by the London Legacy Development Corporation (C02955) to the Global Disability Innovation Hub, University College London. The work was also supported by the EU Future and Emerging Technologies Proactive Programme H2020 (Grant No. 824160: EnTimeMent -https://entimement.dibris.unige.it/). The eSense earbuds used in this work were supported by Nokia Bell Lab, Cambridge, UK.

REFERENCES

[1] K. Doherty and G. Doherty, "The construal of experience in HCI: Understanding self-reports," *International Journal of Human-Computer Studies,* vol. 110, pp. 63-74, 2018.
[2] N. Van Berkel, D. Ferreira, and V. Kostakos, "The experience sampling method on mobile devices," *ACM Computing Surveys (CSUR),* vol. 50, no. 6, pp. 1-40, 2017.
[3] N. B.-B. Tao Bi, Enrico Costanza, Aneesha Singh, "Designing Voice-based Runner Experience Sampling Methods to Build Wearable Runners' Experience Recognition System.," 2022.
[4] T. Bi *et al.*, "Towards Chatbot-Supported Self-Reporting for Increased Reliability and Richness of Ground Truth for Automatic Pain Recognition: Reflections on Long-Distance Runners and People with Chronic Pain," in *Companion Publication of the 2021 International Conference on Multimodal Interaction*, 2021, pp. 43-53.
[5] F. Kawsar, C. Min, A. Mathur, A. Montanari, U. G. Acer, and M. Van den Broeck, "eSense: Open Earable Platform for Human Sensing," in *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 2018: ACM, pp. 371-372.
[6] eSense. "eSense." https://www.esense.io (accessed 13 Sep, 2022).
[7] Nvivo. "Nvivo." https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home (accessed 13 Sep, 2022).
[8] Audacity. "Audacity." https://www.audacityteam.org (accessed 13 Sep, 2022).
[9] Audacity. "Noise Reduction in Audacity." https://manual.audacityteam.org/man/noise_reduction.html (accessed 13 Sep, 2022).
[10] M. Stephen, "Effect of Wind on Sound Transmission," ed. sciencing.com, 2018.
[11] R. Mann *et al.*, "The effect of shoe type and fatigue on strike index and spatiotemporal parameters of running," vol. 42, no. 1, pp. 91-95, 2015.