Prague, July 10–13, 2022
Work-in-Progress, Student paper

Speaker discrimination and classification in breath noises by human listeners

Raphael Werner, Jürgen Trouvain, and Bernd Möbius
Language Science and Technology, Saarland University, Saarbrücken, Germany
{rwerner|trouvain|moebius}@lst.uni-saarland.de

Audible breath noises are frequent companions to speech, occurring roughly every 3 to 4 seconds (Rochet-Capellan & Fuchs, 2013; Kuhlmann & Iwarsson, 2021), and may also be present outside of speech during effortful actions (Trouvain & Truong, 2015). Being a vital function, breathing is arguably less affected by speakers trying to disguise their voice and neural networks have shown promising results on speaker identification based on breath noises (Lu et al, 2020; Zhao, Gao, & Singh, 2017). However, breathing has remained largely untapped for forensic purposes, with few exceptions (eg. Kienast & Glitza, 2003). In this paper we want to investigate the potential that breath noises have for speaker discrimination and classification by human listeners.

We annotated breath noises in dyadic conversations (van Son et al, 2008). For high comparability and since they are most frequent around speech (Lester & Hoit, 2014), we here use 5 audible oral (and probably simultaneously nasal) inhalations each from 6 younger (age range: 20–29; 3m, 3f) and 6 older (age range: 59–65; 3m, 3f) speakers. These noises were then used as stimuli in two tasks: 1) Discrimination task: participants heard 2 breath noises (separated by 500 ms of silence; 14 pairs by participant) and were asked whether they were produced by the same speaker or not. We also recorded participants' confidence on a 5-point Likert scale. 2) Speaker classification task: participants listened to one breath noise at a time (20 noises by participant) and were asked whether the breath noise was produced by a young vs old and male vs female speaker and how confident they were in each of these answers. We recruited and paid 33 speakers (22 f, 10 m, 1 other; age range: 20-71, median: 31), who reported wearing headphones in a quiet environment and having no hearing difficulties, via Prolific (2014) and ran the experiment on Labvanced (Finger et al, 2017).

Preliminary analysis suggests that the discrimination task was answered correctly at 64.3%. In speaker classification, the speaker's age group was correct at a rate of 50.2%, whereas for sex it was 66.7%. The general direction of sex being easier to guess than age here seems to follow the pattern described by Jessen (2007), even though not using speech here and speaker age being a binary decision between two groups. In the further analysis, we will examine what participant or speaker variables contribute to correctness in the discrimination task, as well as look into participant performance by their age and gender.

The findings will have implications for naturalistic synthetic speech and how breath noises there need to be geared to the artificial speaker to be perceived as natural. For forensic purposes, they explore to what extent breath noises may be exploitable for speaker classification and discrimination tasks. It should be borne in mind, however, that all stimuli used here were made under the same recording setup and are thus highly comparable, whereas in real-world forensic applications many factors may complicate comparisons.

References

Finger, H., C. Goeke, D. Diekamp, K. Standvoss, and P. König (2017). LabVanced: A Unified JavaScript Framework for Online Studies. In International Conference on Computational Social Science, 2016–2018.

Jessen, M. (2007). Speaker Classification in Forensic Phonetics and Acoustics. In: Müller, C. (eds) Speaker Classification I. Lecture Notes in Computer Science, vol 4343. Springer, Berlin, Heidelberg.

Kienast, M., & Glitza, F. (2003). Respiratory sounds as an idiosyncratic feature in speaker recognition. ICPhS, 1607–1610.

Kuhlmann, L. L., & Iwarsson, J. (2021). Effects of Speaking Rate on Breathing and Voice Behavior. Journal of Voice.

Lester, R. A., & Hoit, J. D. (2014). Nasal and oral inspiration during natural speech breathing. Journal of Speech, Language, and Hearing Research, 57(3), 734–742.

Lu, L., Liu, L., Hussain, M. J., & Liu, Y. (2020). I Sense You by Breath: Speaker Recognition via Breath Biometrics. IEEE Transactions on Dependable and Secure Computing, 17(2), 306–319.

Prolific. (2014). URL https://www.prolific.co. Accessed: 17/05/2022.

Rochet-Capellan, A., & Fuchs, S. (2013). The interplay of linguistic structure and breathing in German spontaneous speech. Interspeech, 2014–2018.

Trouvain, J., & Truong, K. P. (2015). Prosodic characteristics of read speech before and after treadmill running. Interspeech, 3700–3704.

van Son, R., Wesseling, W., Sanders, E., & van den Heuvel, H. (2008). The IFADV Corpus: a Free Dialog Video Corpus. LREC. 501–508.

Zhao, W., Gao, Y., & Singh, R. (2017). Speaker identification from the sound of the human breath. *arXiv preprint arXiv:1712.00171*.