Speaker discrimination and classification in breath noises by human listeners

Raphael Werner, Jürgen Trouvain, and Bernd Möbius

Language Science and Technology, Saarland University, Saarbrücken

Audible breath noises are frequent companions to speech, occurring roughly every 3 to 4 seconds [1, 2], and may also be present outside of speech during effortful actions [3]. 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 [4, 5]. However, breathing has remained largely untapped for forensic purposes, with few exceptions (e.g. [6]). 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 [7]. For high comparability and since they are most frequent around speech [8], 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 two 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 [9] and ran the experiment on Labvanced [10].

The discrimination task was answered correctly at 64.3 % (sd: 11.8 %), with the lowest results for combinations of different age and same sex. In speaker classification, the speaker's age group was correct at a rate of 50.2 % (sd: 9.1 %), whereas for sex it was 66.7 % (sd: 13.5 %). Confidence did not differ much between tasks or between sex and age in the classification task.

The results in both tasks suggest that sex differences are more perceivable than age differences. This general direction seems to be in line with regular speech [11], even though we used ingressive, unphonated noises only here and speaker age was a very coarse-grained distinction between two separate groups. Perceivable differences by sex but not age may be related to differences in vocal tract length, which differs by sex [12, p. 25-26]. Age differences may thus be audible when comparing children to adults. Although not very high, these numbers suggest that breath noises may be usable for forensic applications to some extent, given that each individual breath noise used here was only 300 to 1000 ms long. Including breathing patterns, rather than just one or two noises, may add to finding speaker-specific characteristics [13].

These findings 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.

doi:10.11576/pundp2022-1050

References

[1] Amélie Rochet-Capellan and Susanne Fuchs. The interplay of linguistic structure and breathing in German spontaneous speech. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, pages 2014–2018, 2013.

[2] Laura Lund Kuhlmann and Jenny Iwarsson. Effects of Speaking Rate on Breathing and Voice Behavior. *Journal of Voice*, 2021.

[3] Jürgen Trouvain and Khiet P Truong. Prosodic characteristics of read speech before and after treadmill running. In *Interspeech 2015*, pages 3700–3704, ISCA, 2015. ISCA.

[4] I Sense You by Breath: Speaker Recognition via Breath Biometrics. *IEEE Transactions on Dependable and Secure Computing*, 17(2):306–319, 2020.

[5] Wenbo Zhao, Yang Gao, and Rita Singh. Speaker identification from the sound of the human breath. *CoRR*, abs/1712.00171, 2017.

[6] Miriam Kienast and Florian Glitza. Respiratory sounds as an idiosyncratic feature in speaker recognition. In *Proceedings of 15th ICPhS*, pages 1607–1610, Barcelona, 2003.

[7] R. J.J.H. Van Son, Wieneke Wesseling, Eric Sanders, and Henk Van Den Heuvel. The IFADV corpus: A free dialog video corpus. *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*, 2(1):501–508, 2008.

[8] Rosemary A. Lester and Jeannette D. Hoit. Nasal and oral inspiration during natural speech breathing. *Journal of Speech, Language, and Hearing Research*, 57(3):734–742, 2014.

[9] Prolific, 2014. Accessed: 17/05/2022.

[10] Holger Finger, Caspar Goeke, Dorena Diekamp, Kai Standvoß, and Peter König. LabVanced: A Unified JavaScript Framework for Online Studies. In *2017 International Conference on Computational Social Science IC2S2*, 2017.

[11] Michael Jessen. Speaker classification in forensic phonetics and acoustics. In Christian Müller, editor, *Speaker Classification I: Fundamentals, Features, and Methods*, pages 180–204. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[12] Kenneth N Stevens. *Acoustic phonetics*, volume 30. MIT press, 2000.

[13] Hélène Serré, Marion Dohen, Susanne Fuchs, Silvain Gerber, and Amélie Rochet-Capellan. Speech breathing: variable but individual over time and according to limb movements. *Annals of the New York Academy of Sciences*, 1505(1):142–155, 2021.

doi:10.11576/pundp2022-1050