Breath noise perception – a pilot study on airway usage

Raphael Werner, Jürgen Trouvain, Beeke Muhlack and Bernd Möbius

In this study, we aim to find out how well listeners can discriminate between different types of breath noises in speech concerning air flow direction (inhalation or exhalation) and airway usage (oral or nasal). This experiment is based on audio alone which can be useful when looking at respiration in detail, e.g. in [1-3], and for its acoustic analysis, e.g. in forensics or automatic breath detectors.

We annotated breath noises for their type in an audio-visual corpus [4] so that mouth opening was detectable in the videos as a cue for oral contribution. Agreement between the two annotators on a subset comprising 20 % of the data was 92 %; none of the dubious cases were used. Six frequent types of breath noises were chosen: for exhalation: oral, nasal; for inhalation: oral, nasal, oral followed by nasal, nasal followed by oral. 48 samples were used as stimuli, balanced by type and two conditions: no context (only the breath noise itself) and context (including 1 sec of audio before and after). Audio was presented as a web-based listening test to two groups of participants: 8 lay persons and 8 phonetics experts.

Initial results show an overall correct identification for 73.6 % of the stimuli: 76.8 % with vs. 70.3 % without context. There was hardly any difference between experts (74.0 %) and lay persons (73.2 %). Nasal exhalations (72.7 %), as well as oral (72.7 %), nasal-oral (75.0 %), and oral-nasal (67.2 %) inhalations all appeared similarly difficult to recognize, whereas oral exhalations (59.4 %) and even more so nasal inhalations (94.5 %) differed.

For both experts and lay persons, audio-based classification is more or less reliably possible, though with differences for different breath noise types. Context appears to help which might work on a smaller (e.g. nasal inhalations after/before nasal sounds) or larger (e.g. exhalations often appearing outside of fluent speech) scale.

[1] Trouvain, J., & Belz, M. (2019). Zur Annotation nicht-verbaler Vokalisierungen in Korpora gesprochener Sprache. ESSV 2019, 280–287.

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

[3] Scobbie, J. M., Schaeffler, S., & Mennen, I. (2011). Audible aspects of speech preparation. ICPhS XVII, 1782–1785.

[4] van Son, R. J. J. H., Wesseling, W., Sanders, E., & Van Den Heuvel, H. (2008). 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.