Perceptual categorization of breath noises in speech pauses

Raphael Werner, Jürgen Trouvain, Beeke Muhlack, Bernd Möbius
{rwerner|trouvain|muhlack|moebius}@lst.uni-saarland.de

1 Introduction

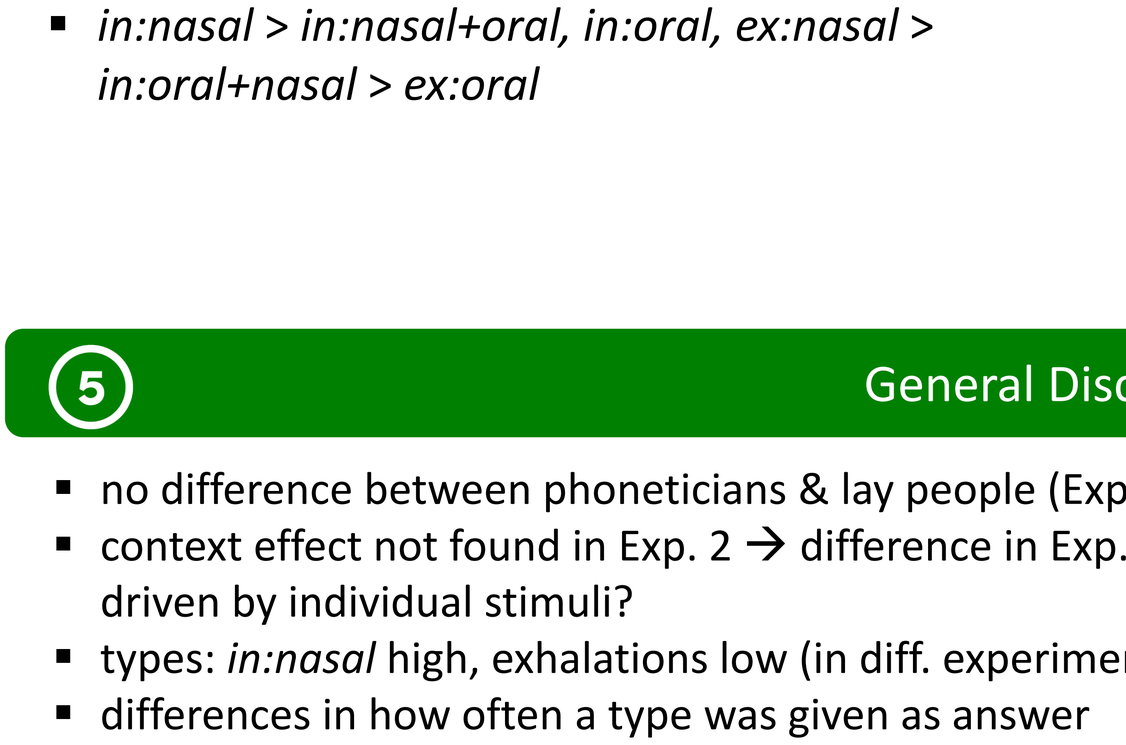
- breathing possible in various ways and combinations
  - air flow direction (in- vs exhalation)
  - airway (oral, nasal, simultaneous oral-nasal, alternations beginning with oral or nasal)
- breath noise categorization by audio relevant for investigating respiration in detail [1-3], annotation, or their acoustic analysis
- research questions:
  - how reliable is the audio categorization of breath noises?
  - does context (+1sec before & after) help?
  - are phoneticians better than lay people?
  - are there differences by breath noise category?

2 Methods

- 20 speakers (10m, 10f) from Dutch audio-visual corpus [4]
  → mouth opening as visual cue for oral contribution
- 812 breath noises annotated by 2 raters (inter-rater agreement on 20% subset ≈ 92%, Cohen's κ = .88)
- 6 frequent types chosen:
  - exhalation: oral, nasal
  - inhalation: oral, nasal, oral+nasal, nasal+oral

3 Experiment 1

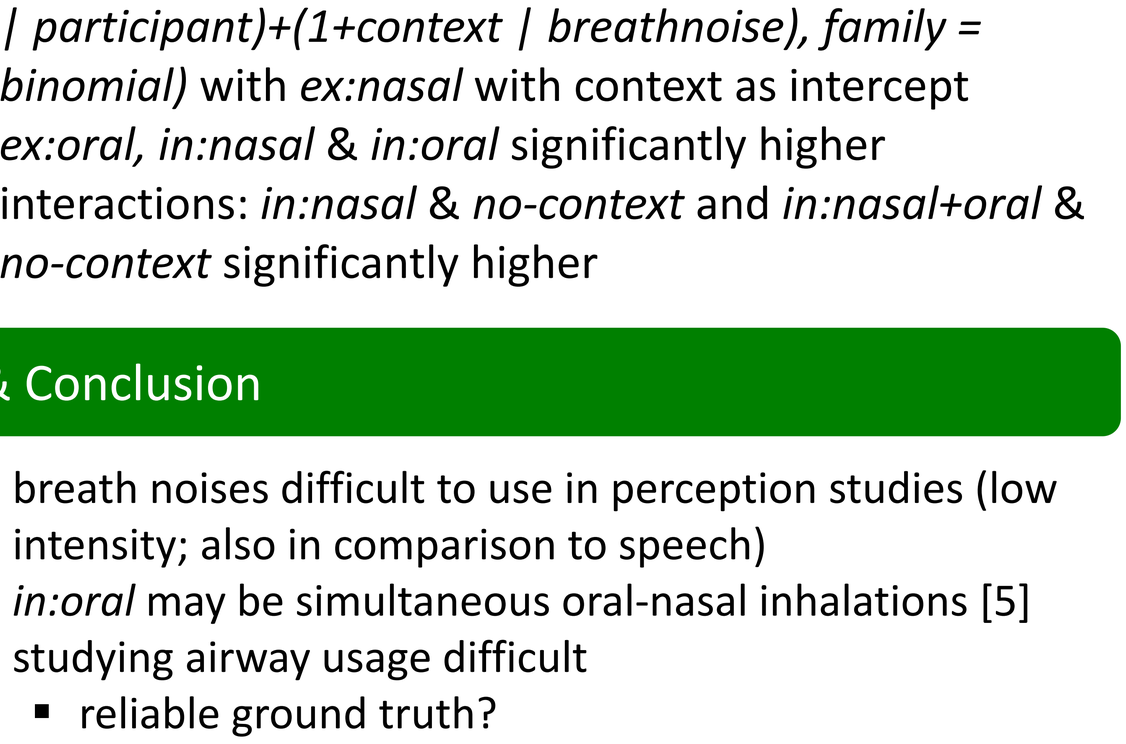
- 2 conditions (with/without 1 sec context); randomly selected 4 noises per type & condition
- 48 individual stimuli assessed by 8 phoneticians & 8 lay people via Labvanced



- overall correctly identified: 73.6 %
- with context (76.8%) > without context (70.3 %)
- phoneticians (74.0 %) ≈ lay people (73.2 %)
- *in:nasal* > *in:nasal+oral, in:oral, ex:nasal* > *in:oral+nasal* > *ex:oral*

4 Experiment 2

- stimuli matched for context→ 2 lists of 24 breath noises to present via Labvanced
- 80 native German participants via Prolific; mean age 34 years (range 18–72)



- overall correctly identified: 65.8 %
- with context (66.7 %) ≈ without context (65 %)
- *glmer(correct ~ breathtype * context + (1+breathtype | participant)+(1+context | breathnoise), family = binomial)* with *ex:nasal* with context as intercept
- *ex:oral, in:nasal* & *in:oral* significantly higher
- interactions: *in:nasal* & *no-context* and *in:nasal+oral* & *no-context* significantly higher

5 General Discussion & Conclusion

- no difference between phoneticians & lay people (Exp. 1)
- context effect not found in Exp. 2 → difference in Exp. 1 driven by individual stimuli?
- types: *in:nasal* high, exhalations low (in diff. experiments)
- differences in how often a type was given as answer (regardless of stimulus)
- interaction: no context beneficial for 2 types
- breath noises difficult to use in perception studies (low intensity; also in comparison to speech)
- *in:oral* may be simultaneous oral-nasal inhalations [5]
- studying airway usage difficult
  - reliable ground truth?
  - non-invasive, non-influential measurement?
- overall rate of around 2/3 correct→ reliable/usable?

6 References

[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 et al. (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.
[5] Lester, R. A. & Hoit, J. D. (2014). Nasal and oral inspiration during natural speech breathing, J. Speech, Lang. Hear. Res., vol. 57, no. 3, 734–742.