Speaker discrimination and classification in breath noises by human listeners

Deutsche Forschungsgemeinschaft

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

UNIVERSITÄT DES SAARLANDES

1 Introduction

- audible breathing frequent around speech [1, 2] or during effortful actions [3]
- as vital function, perhaps less affected by disguising voice
- breath rarely used for forensic purposes (e.g. [4, 5])
- speaker identification by neural networks looks promising [6, 7]

- research questions:
  - (A) how well can listeners discriminate between same vs different breathers?
  - (B) how well can listeners guess a breather's age (young vs old) and sex (male vs female)?

2 Methods

- breath noises annotated in conversations [8]
- 5 oral(+nasal) inhalations each from 6 young (20-29 yrs; 3f, 3m) and 6 old speakers (age: 59-65 yrs; 3f, 3m)
- 33 participants (22f, 10m, 1 other; age: 20-71 yrs, median: 31 yrs) via Prolific [9] and Labvanced [10]

(A) discrimination task: 2 breath noises, separated by 500 ms silence → same or different speaker? how confident (1-5)?

(B) classification task: 1 breath noise → speaker young/old? male/female? how confident each (1-5)?

3 Results

(A) discrimination task

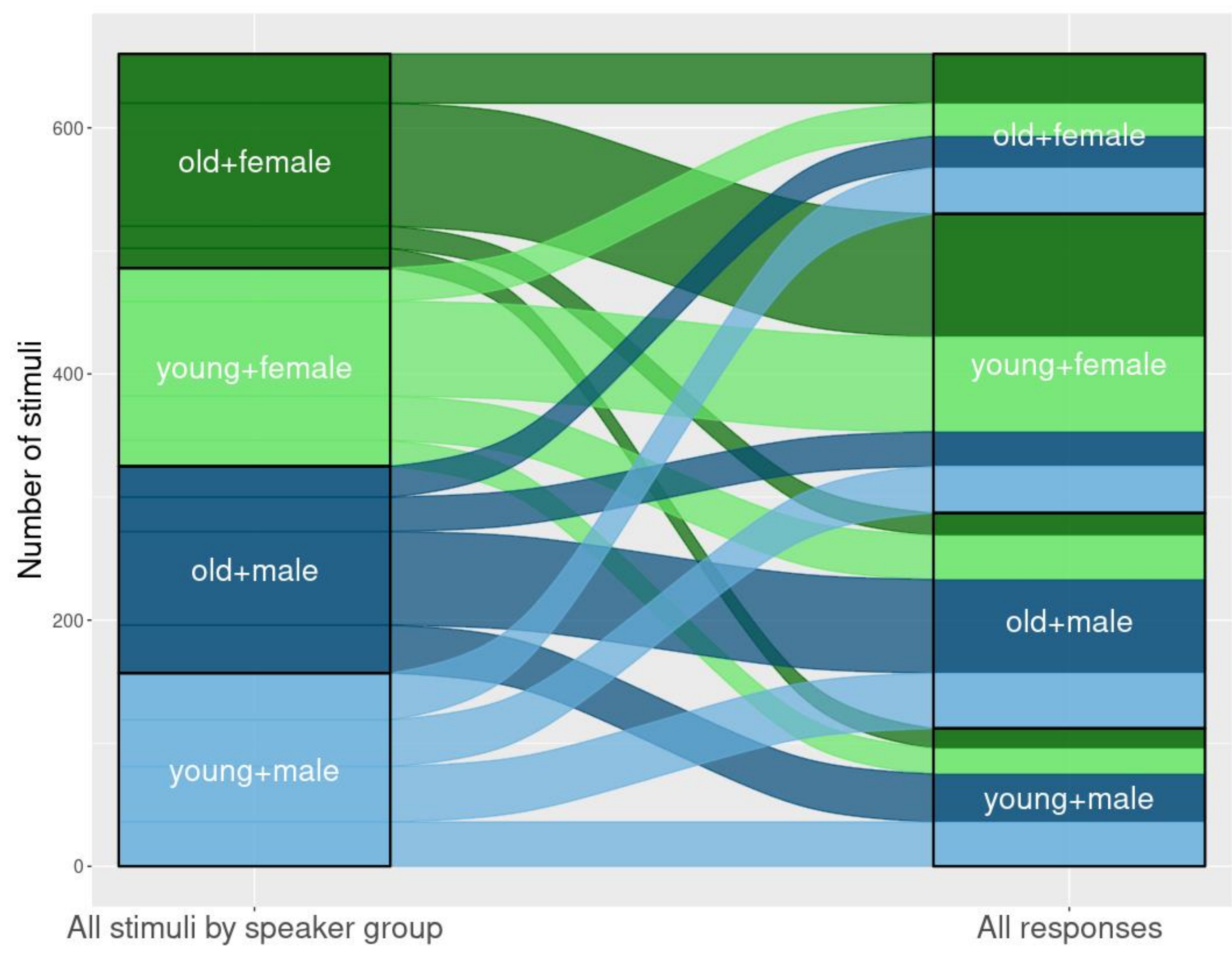
- **overall** correctness rate: mean = **64.3%** (sd: 11.8%)
- confidence rating: mean = 3.5 (sd: 0.76)
- sex differences seem more perceivable than age differences
- different age + same sex even far below chance
- young, female speakers stand out

		sex			
		same_male	same_female	different	total
age	same_old	79.5% (73)	76.8% (69)	60.0% (35)	74.6% (177)
	same_young	73.1% (67)	53.1% (64)	65.0% (40)	63.7% (171)
	different	31.8% (22)	35.3% (34)	63.8% (58)	49.1% (114)
	total	70.4% (162)	59.3% (167)	63.2% (133)	64.3% (462)

Table: Correctness rate by speaker sex and age in percent. Numbers in brackets indicate number of stimuli per cell.

(B) classification task

- overall correctness rate:
  - **age**: mean = **50.2%** (sd: 9.1%); confid. = 3.0 (sd: 0.75)
  - **sex:** mean = **66.7%** (sd: 13.5%); confid. = 3.2 (sd: 0.77)



4 Discussion and Conclusion

- speaker **discrimination** possible, but not with high accuracy
- **classification**: sex > age (in line with findings for regular speech from [11])
- only binary distinctions for two categories here
- confounding factors: biological vs chronological age? height/weight?
- implications for using breath noises in synthetic speech
- breath noises relevant in real-world forensic applications (e.g. rape, black box)

5 References

[1] Rochet-Capellan, A., & Fuchs, S. (2013). The interplay of linguistic structure and breathing in German spontaneous speech. Interspeech, 2014–2018. **[2]** Kuhlmann, L. L., & Iwarsson, J. (2021). Effects of speaking rate on breathing and voice behavior. Journal of Voice. **[3]** Trouvain, J., & Truong, K. P. (2015). Prosodic characteristics of read speech before and after treadmill running. Interspeech, 3700–3704. **[4]** Kienast, M., & Glitza, F. (2003). Respiratory sounds as an idiosyncratic feature in speaker recognition. ICPhS, 1607–1610. **[5]** Braun, A. (2017). Nonverbal vocalisations – a forensic phonetic perspective. Laughter and other non-verbal vocalisations workshop, pp. 19–23. 2020. **[6]** Lu, L. et al. (2020). I sense you by breath: speaker recognition via breath biometrics. IEEE Transactions on Dependable and Secure Computing, 17(2), 306–319. **[7]** Zhao, W., Gao, Y., & Singh, R. (2017). Speaker identification from the sound of the human breath. **[8]** van Son, R. et al. (2008). The IFADV corpus: a free dialog video corpus. LREC. 501–508. **[9]** Prolific. (2014). URL https://www.prolific.co. Accessed: 17/05/2022. **[10]** Finger, H. et al. (2017). LabVanced: a unified JavaScript framework for online studies. In International Conference on Computational Social Science, 2016–2018. **[11]** 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.