

The impact of social information on VOT shadowing by nonbinary speakers



Jack Rechsteiner & Betsy Sneller
Michigan State University

Introduction

- Social information can impact the degree to which one speaker phonetically converges with another speaker [1, 2, 3, 4, 5].
- Nonbinary speakers alter their speech in queer vs. non-queer settings, where there is a threat of being misgendered [6].
- Here, we investigate whether nonbinary speakers' convergence toward extended voice onset time (VOT) in word-initial English /p, t, k/ is impacted by whether they believe they are listening to another nonbinary speaker or to a cis speaker.
- We predict that nonbinary speakers will converge most strongly towards a nonbinary model talker.**

Methods

Participants

15 American English speakers (ages 18-35) who reported they are nonbinary, born in the US, and currently live in the US.

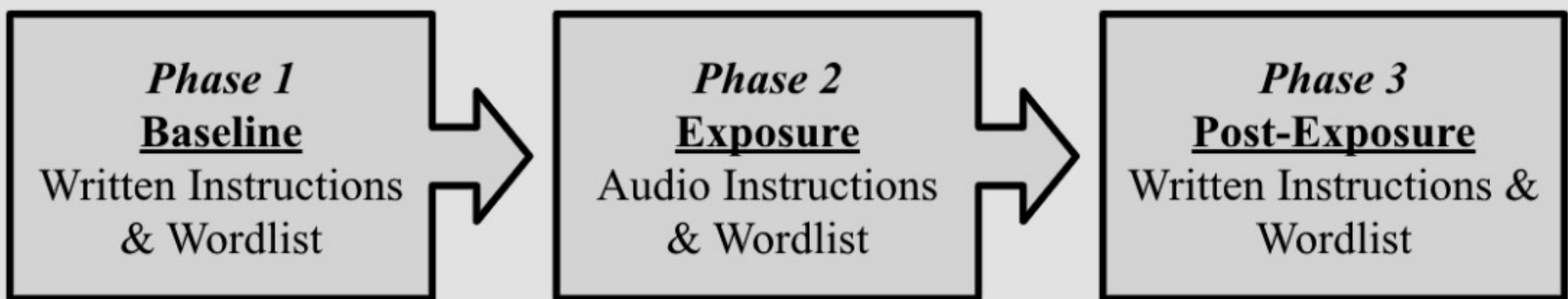
Stimuli

The stimuli consisted of 54 words — 40 target words and 14 filler words. All stimuli were bisyllabic, stress-initial words with a frequency between 1 and 25 per million based on SUBTLEXUS scores [7]. This stimuli set is consistent with stimuli used in previous studies on extended VOT convergence [2, 3, 8]. Mean frequencies per million shown below:

Initial Stop	Mean FPM	Example word
/p/	9.87	pollen
/t/	13.21	timber
/k/	10.78	cabin

Procedure

Each participant was assigned to 1 of 3 conditions (Cis, Nonbinary, Neutral) and took part in 3 phases of a shadowing task where they recorded themselves saying the given word within a carrier phrase.



In the Exposure Phase, participants were assigned to 1 of 3 conditions, differing by what social information was given by the model talker: **Nonbinary** (“My name is Sam and my pronouns are they/them”), **Cis** (“My name is Grant and my pronouns are he/him”), and **Neutral** (no social information given).

Aside from gender identity information, the recordings were from the same model speaker and were identical in each condition. Participants were distributed evenly across the 3 conditions (5 participants in each condition).

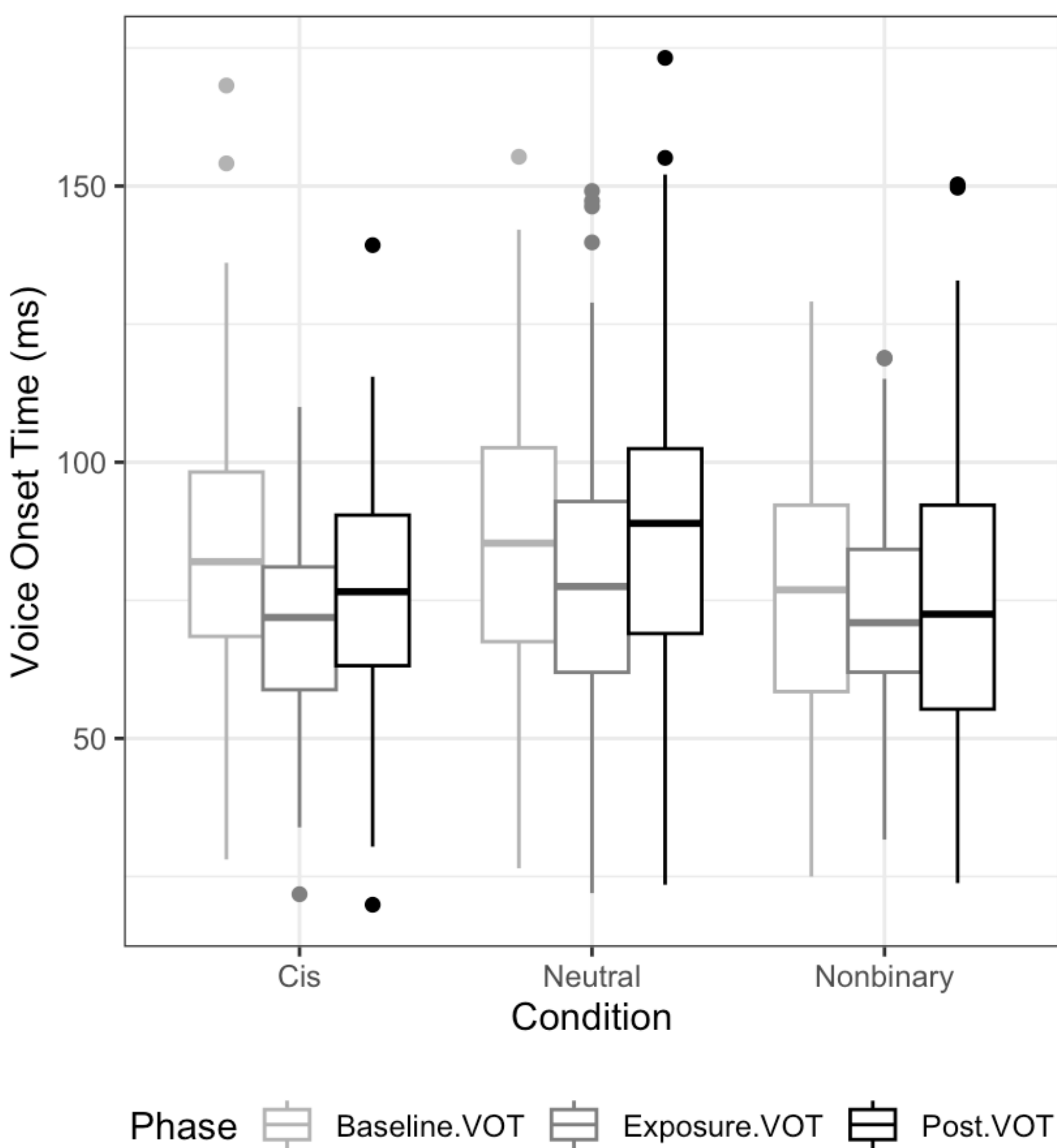


Scan the QR code above to read the full paper.

Results

Unexpectedly, all conditions saw a **decrease** in participant VOT values during the Exposure Phase compared to their Baseline Phase, suggesting divergence from the model talker.

In line with our hypothesis, nonbinary speakers diverge the *least* from a nonbinary model talker.



Linear mixed-effects model:

VOT ~ Phase*Condition + Stop + Vowel + (1|Speaker) + (1|Word)

Fixed Effect	Estimate	P Value
Neutral Baseline VOT	77.47	<.001***
(Neutral) Exposure	-8.07	<.0001***
(Neutral) Post	1.94	.21
Cis (Baseline)	-2.81	.68
Cis : Exposure	-4.94	.03*
Cis : Post	-9.52	<.001***
Nonbinary (Baseline)	-11.12	.12
Nonbinary : Exposure	5.86	.008**
Nonbinary : Post	-2.94	.18
Initial Stop /p/	-9.32	.01*
Initial Stop /t/	5.53	.19
Vowel Height Low	18.91	<.001***
Vowel Height Mid	7.11	.08

- Average VOT value for Baseline phase of the Neutral condition was 77.47 ms.
- Exposure phase of the Neutral condition shows participants significantly ($p < 0.001$) diverging from the model talker by decreasing VOT by 8.07 ms.
- Exposure phase of the Cis condition shows a marginally significant effect ($p = 0.03$) for participants to **diverge even more in the Cis Exposure phase** than in the Neutral Exposure phase (an additional 4.94 ms shorter).
- Post phase of the Cis condition shows participants **maintaining their divergence into the post-exposure phase for the Cis condition** ($p < 0.001$, -9.52 ms).
- Exposure phase of the Nonbinary condition ($p = 0.008$) shows participants in the Nonbinary condition still diverging (-8.07 main effect + 5.86 interaction effect = -2.21 ms), but **diverging significantly less than the Neutral or Cis conditions**.
- There are also significant main effects of initial stop /p/ showing shorter VOT ($p = 0.1$, -9.32 ms), and low vowels showing longer VOT ($p < 0.001$, 18.91 ms).

Discussion

- Patterns of consistent divergence away from a model talker highlight that phonetic imitation is mediated by social factors [4, 5]. For example, Babel [4] found that male participants who rated a model talker as attractive were more likely to diverge from that talker, positing that these participants “were, perhaps, *socially threatened* and distanced themselves in response to the threat” (emphasis ours).
- In our study, the difference in divergence across conditions also shows the influence of social factors.
 - Nonbinary participants diverged **most** from a Cis talker (-9.52 ms, $p < 0.001$).
 - We posit that nonbinary participants **linguistically distanced themselves** from a model talker due to an interpreted **social threat**, such as the threat of being misgendered [6].
 - VOT values from the Exposure Phase diverged the least in the Nonbinary Condition (5.86 ms, $p = 0.008$), suggesting that **nonbinary participants align their speech most closely to a nonbinary model talker**

Conclusion

- Results suggest that even in low-interaction virtual settings, **being in an explicitly queer context enables nonbinary speakers to pattern more like another nonbinary speaker** than like a cis-identified speaker.
- Pardo [5] noted that phonetic imitation “is subject to situational constraints that influence the direction and magnitude of phonetic convergence”, and this is precisely what our findings show. Different situational contexts — in this case, whether nonbinary participants have entered an explicitly queer virtual environment or an explicitly heteronormative one — impact phonetic imitation.
- These findings furthermore align with previous work which argued that in conversational speech in queer contexts, nonbinary speakers' pattern more like each other regardless of sex assigned at birth, **effectively creating a distinct nonbinary speech community** [6, 9].

References

- [1] S. D. Goldinger, “Echoes of echoes? an episodic theory of lexical access,” *Psychological Review*, vol. 105, no. 2, pp. 251–279, 1998.
- [2] K. Shockley, L. Sabadini, and C. A. Fowler, “Imitation in shadowing words,” *Perception & Psychophysics*, vol. 66, no. 3, pp. 422–429, 2004.
- [3] K. Nielsen, “Specificity and abstractness of VOT imitation,” *Journal of Phonetics*, vol. 39, no. 2, pp. 132–142, Apr. 2011.
- [4] M. Babel, “Evidence for phonetic and social selectivity in spontaneous phonetic imitation,” *Journal of Phonetics*, vol. 40, no. 1, pp. 177–189, 2012.
- [5] J. S. Pardo, “On phonetic convergence during conversational interaction,” *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.
- [6] C. Gratton, “Resisting the Gender Binary: The Use of (ING) in the Construction of Non-binary Transgender Identities,” *University of Pennsylvania Working Papers in Linguistics*, vol. 22, no. 2, 2016.
- [7] M. Brysbaert and B. New, “Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English,” *Behavior Research Methods*, vol. 41, no. 4, pp. 977–990, Nov. 2009.
- [8] J. Schertz, E. K. Johnson, and M. Paquette-Smith, “The independent contribution of voice onset time to perceptual metrics of convergence,” *JASA Express Letters*, vol. 1, no. 4, p. 045205, Apr. 2021.
- [9] J. Rechsteiner and B. Sneller, “Nonbinary speakers' use of (ING) across gender-related topics,” 2021, NAWAV 49.