THE IMPACT OF SOCIAL INFORMATION ON VOT SHADOWING BY NONBINARY SPEAKERS

Jack Rechsteiner & Betsy Sneller

Michigan State University
rechste4@msu.edu, sneller7@msu.edu

ABSTRACT

Social information can impact the degree to which one speaker phonetically converges with another speaker. There is also evidence that nonbinary speakers alter their speech due to their social environment, specifically in environments where there is a threat of being misgendered. In this paper, 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 tested 15 speakers in an online VOT shadowing input-driven elicitation task, and we found that nonbinary speakers show statistically significant greater divergence away from the cis-labeled voice than in other conditions. These results suggest that the threat of being misgendered is a primary motivation for nonbinary speakers shifting their linguistic productions in differing social contexts.

Keywords: sociophonetics, nonbinary, phonetic imitation, social information, voice onset time

1. INTRODUCTION

Existing research on individuals who are not cisgender — or 'cis', referring to those whose gender identity matches their sex-assigned-at-birth — has largely focused on the experiences of trans people with binary trans identities [1]. However, the amount of research on speakers with gender identities that are outside of the gender binary is significantly smaller [2, 3, 4]. Nonbinary individuals present direct opposition to gender binaries, in that they may identify as being somewhere along the spectrum between male and female or outside of the male-female dichotomy entirely. The experiences of nonbinary individuals have been shown to be different from the experiences of binary-gendered individuals, such as marginalization in both cis and LGBT communities in addition to the unique challenges of being gender non-conforming in spaces that tend to understand gender as a binary framework [2, 1].

Gratton [5, 6] observed nonbinary participants varying their linguistic patterns in queer contexts compared to cisgender contexts, and argued this was motivated by the desire to counteract the possibility of their gender being assumed incorrectly. The present study aims to build upon these findings by analyzing the results of 15 nonbinary American English speakers who participated in an online phonetic imitation shadowing task. This experimental paradigm was chosen in order to investigate the degree to which nonbinary speakers are influenced by socially salient identities, even in minimally interactive conditions. We hypothesize that participants who are exposed to a model speaker stated to share a queer identity with them will show higher rates of convergence than participants who are exposed to a model speaker who does not share a queer identity with them.

1.1. Phonetic convergence

Phonetic convergence is a process whereby a speaker takes on acoustic-phonetic traits that are present in the speech of a person they are interacting with. Convergence falls within the broader category of linguistic accommodation, which has been argued to be "motivated by a desire to affiliate with or decrease social distance to a fellow interactant" as well as "underscoring common social identities" [7]. While phonetic convergence is at least partly socially facilitated [8, 9] speakers also exhibit phonetic convergence in minimally social laboratory settings [10, 11, 12] as well as in cooperative and conversational situations [13, 14].

1.2. Voice onset time

The specific phonetic variable of interest in the current study is voice onset time (VOT), which is defined as the length of time between the release of a stop consonant and the onset of voicing for the following vowel. For the purpose of this study, VOT is equivalent to a practical measurement of the aspiration of voiceless stops. VOT has been well documented as a phonetic object that

produces convergence [12, 15, 16]. Nielsen [15] found speakers extending their VOT productions for stress-initial words beginning with a voiceless stop after exposure to recordings of a model speaker with artificially extended VOT in a non-shadowing elicitation task. Extended VOT was chosen as our variable of interest because it is not an explicit stereotype of gender (in the sense of Labov 1972's *indicators, markers*, and *stereotypes* [17]), extended VOT stimuli are easy to artificially create through acoustic manipulation, and extended VOT has no phonological perception consequences for voiceless stops in English [15].

2. METHODS

2.1. Participants

The participants in this study included 15 American English speakers (ages 18-35, mean age 27) who identified as nonbinary and reported that they were born and currently live in the United States. Participants were recruited through the researchers' social media networks. Participating in the study took roughly 15 minutes from beginning to end, and participants were compensated for their participation.

2.2. Stimuli

The stimuli consisted of 54 words — 40 target words and 14 filler words. All of the words in the stimuli are bisyllabic, stress-initial words with a frequency between 1 and 25 per million based on frequency scores provided by the SUBTLEXUS database [18]. Low frequency words were chosen because previous research has shown phonetic properties in low frequency words to show a higher degree of convergence than high frequency words [10]. For the target words, 16 have word-initial /p/, 16 have word-initial /k/, and 8 have word-initial /t/; no target words have initial onset clusters. All 14 filler words begin with vowels. This stimuli set is consistent with the stimuli used in previous studies on extended VOT convergence [12, 15, 16]. Mean frequencies for the words in each category of word-initial stop can be seen in Table 1.

The model speaker was an American English speaker who was determined to sound appropriately gender-ambiguous to listeners via a pre-experiment norming study. The model speaker provided recordings of the 54 stimuli words, as well as the audio instruction portion of the experiment. The original VOT of initial consonants was measured and then extended using the Duration Tier in Praat's manipulation features [19] to create VOTs that were, on average, 102% longer than the original VOT. This method was chosen in order to avoid auditory aberrations, such as aperiodic bursts, that can occur when manipulating VOT through other means.

Initial stop	Mean FPM	Example word
/p/	9.87	pollen
/t/	13.21	timer
/k/	10.78	cabin

Table 1: Mean frequency per million (FPM) of target words by word-initial stop with examples of words used.

2.3. Procedure

The experiment consists of a shadowing input-driven elicitation task where each participant is assigned to 1 of 3 conditions for a between-subject experiment design. In this experimental paradigm, words are presented to participants before, during, and after exposure to a model speaker and participants record themselves speaking the word aloud. The experiment was built with and administered online using PsychoPy [20]. After giving informed consent, participants completed a demographic survey to collect information on their age, gender identity, residential history, race/ethnicity, and education. Participants then took part in 3 phases of the shadowing task where they recorded themselves saying the given word within the carrier phrase, "The word is ____."

Phase 1 (Baseline Phase) elicited participants' baseline productions by presenting written instructions and words on the screen without any auditory exposure. Phase 2 (Exposure Phase) presented participants with audio instructions and words read aloud by the model speaker. Phase 3 (Post-exposure Phase) again presented participants with written words with no accompanying audio. The order in which the words were presented in each phase was randomised for all participants.

In the Exposure Phase, participants were given auditory instructions from 1 of 3 conditions. In the Nonbinary Condition, the model speaker begins by explicitly identifying themself as nonbinary ("My name is Sam. I am nonbinary and my pronouns are they/them"). In the Neutral Condition, the model speaker does not give any information about their gender ("My name is Sam"). In the Cis Condition, the speaker explicitly identifies themself as cis ("My name is Grant and my pronouns are he/him").

Aside from this introductory gender identity information, the recordings for the model speaker

were identical in each condition. The recordings were from the same model speaker in each condition, and the pre-experiment norming study on gender ambiguity of the model speaker aimed to mitigate effects that would cause participants to assume the gender of the speaker in the Neutral Condition. The structure of these conditions was motivated by the hypothesis that nonbinary speakers are more likely to converge with a model speaker they perceive as nonbinary, as suggested by the results from Gratton [5, 6] which showed nonbinary participants were more likely to pattern together in queer spaces than in non-queer spaces. Participants were distributed evenly across the 3 conditions, resulting in 5 participants for each condition.

3. RESULTS

Following Nielsen [15], the VOT of participant responses was measured in Praat [19]. Recordings of the shadowing task were transcribed orthographically and force aligned using the Montreal Forced Aligner [21] via DARLA [22]. Measuring the VOT of target words in Praat was done manually, assisted by the get_vot Praat script [23].

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 (Fig. 1). This may be the result of participants using a hyperarticulated "citation style" in the Baseline phase, and becoming more familiar with the task in subsequent phases, or it could be the result of social divergence. Either way, our focus here is not on the direction of effect, but rather on the differences in degree of divergence across the three social conditions. These results were analyzed using a linear mixed-effects model in RStudio [24], with VOT as the dependent variable and an interaction term between the fixed effects of Experiment Phase (Baseline, Exposure, and Post-Exposure) and Condition (Neutral, Cis, and Nonbinary), fixed effects for the initial stop (p, t, k) and the height of the following vowel (low, mid, high). Random intercepts were included for speaker and word. Table 2 shows an overview of the statistics of this model. The model formula used was:

$$(1)\quad lmer(VOT \sim Phase * Condition + Initial\,stop + Vowel\,height + (1|Speaker) + (1|Word))$$

Here, we discuss the significant results. The reference level shows that the average VOT value for participants in the Baseline Phase of the Neutral

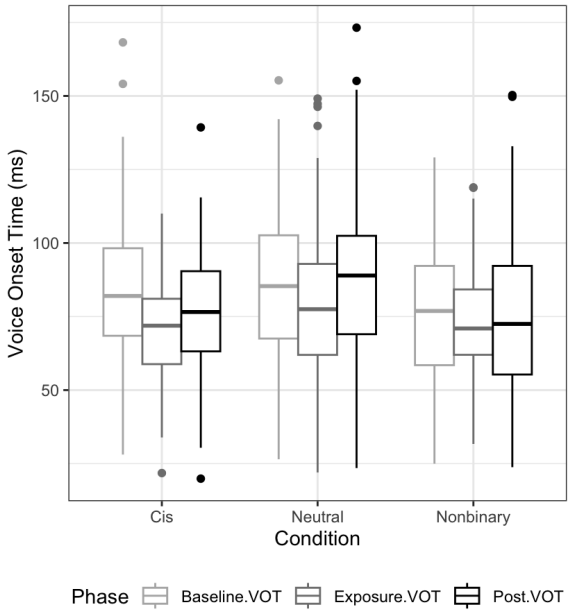

Figure 1: VOT values across the 3 conditions.

Fixed Effect	Estimate	P value
Neutral Baseline VOT	**77.47**	**<.001*** **
(Neutral) Exposure	**-8.07**	**<.001*** **
(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

Table 2: Results of the mixed effects model.

condition was 77.47 ms. We see a significant main effect ($p < 0.001$) of Exposure phase for the Neutral condition, indicating that participant VOTs diverged from the model talker by *decreasing* by 8.07 ms. The interaction between Cis condition and Exposure phase shows a marginally significant effect ($p = 0.03$), showing participants diverging even more in the Cis condition Exposure phase than in the Neutral condition Exposure phase (an additional 4.94 ms shorter, on top of the main effect of Exposure phase). The interaction between Cis condition and Post phase shows participants in the Cis condition maintaining their divergence ($p <$

0.001) in the Post exposure phase, meaning that their divergence from the model talker persisted even beyond immediate exposure. The interaction between Nonbinary condition and Exposure phase ($p = 0.008$) shows participants in the Nonbinary condition still diverging (-8.07 main effect + 5.86 interaction effect = -2.21 ms), but significantly *less* than participants in the Neutral or Cis conditions. Finally, we find expected significant main effects of initial stop, with /p/ showing significantly shorter VOT ($p = 0.1$, 9.32 ms), and low vowels showing significantly longer VOT ($p < 0.001$, 18.91 ms).

4. DISCUSSION

This study examined VOT imitation effects in 15 American English speakers across 3 different experimental conditions with the prediction that nonbinary speakers would show stronger convergence when they were told the model speaker is also nonbinary (Nonbinary Condition) as compared to the other conditions. This prediction was based on previous observations that suggest the threat of being misgendered is a primary motivation for nonbinary speakers shifting their linguistic productions in differing social contexts [5, 6].

The results showed a surprising tendency for participants in all 3 conditions to diverge from, rather than converge with, the model speaker's VOT. Patterns of consistent divergence away from a model talker, like those seen in this study, highlight that phonetic imitation is not simply an automatic process, but instead mediated by social factors [8, 13, 14]. For example, Babel [8] found that male participants who rated a model talker as attractive were more likely to diverge from that talker's production. Babel posits that these participants "were, perhaps, *socially threatened* and distanced themselves in response to the threat" (emphasis ours). In our case, the difference in divergence across conditions also shows the strong influence of social factors. We found that nonbinary participants diverged the *most* in the Cis Condition (-9.52 ms, $p < 0.001$). We posit that nonbinary participants interpreted a social threat associated with a cis model talker, such as the threat of being misgendered [5, 25], which was strong enough to motivate participants to linguistically distance themselves from a cis-identified talker.

Additionally, 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 model talker when they are explicitly identified as sharing a nonbinary identity. We interpret this that participants who are in an explicitly queer virtual setting, even a very low-interaction one, converge towards a shared nonbinary speech norm.

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 [5, 26].

5. CONCLUSION

This study aimed to investigate the impact that model talker gender identity has on the direction and degree to which nonbinary speakers converge in VOT. We found that compared to a model talker who is unlabeled for gender identity, a nonbinary model talker resulted in significantly less divergence for nonbinary participants. We additionally found that a cis-labeled model talker resulted in significantly more divergence for nonbinary participants. These 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.

Previous phonetic imitation studies have shown that speaker gender does not have a consistent, significant effect on imitation [13, 27, 28]. This is not to say, however, that gender does not matter for imitation or convergence. Pardo [13] 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.

6. REFERENCES

[1] A. E. Goldberg and K. A. Kuvalanka, "Navigating identity development and community belonging when "there are only two boxes to check": An exploratory study of nonbinary trans college students," *Journal of LGBT Youth*, vol. 15, no. 2, pp. 106–131, Apr. 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/19361653.2018.1429979

[2] N. J. Bradford, G. N. Rider, J. M. Catalpa, Q. J. Morrow, D. R. Berg, K. G. Spencer, and J. K. McGuire, "Creating gender: A thematic analysis of genderqueer narratives," *International Journal of Transgenderism*, vol. 20, no. 2-3, pp. 155–168, Jul. 2019. [Online].

Available: https://www.tandfonline.com/doi/full/10.1080/15532739.2018.1474516

[3] A. Garmpi, "The Discursive Construction and Performance of Non-Binary Identity," Master's thesis, The University of Edinburgh, 2020.

[4] J. Jones, "Authentic Self, Incongruent Acoustics: A Corpus-Based Sociophonetic Analysis of Nonbinary Speech," Ph.D. dissertation, University of Canterbury, 2022.

[5] 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. [Online]. Available: https://repository.upenn.edu/pwpl/vol22/iss2/7

[6] ——, "Non-binary identity construction and intraspeaker variation," in *The 91st Annual Meeting of the Linguistic Society of America, University of Texas at Austin*, 2017. [Online]. Available: https://academia.edu/30897570/Non_binary_Identity_Construction_and_Intraspeaker_Variation

[7] J. Gasiorek, H. Giles, and J. Soliz, "Accommodating new vistas," *Language & Communication*, vol. 41, pp. 1–5, 2015.

[8] M. Babel, "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," *Journal of Phonetics*, vol. 40, no. 1, pp. 177–189, 2012.

[9] N. Coupland, "Accommodation at work: Some phonological data and their implications," *International Journal of the Sociology of Language*, vol. 39, no. 2, pp. 49–70, 1984.

[10] S. D. Goldinger, "Echoes of echoes? an episodic theory of lexical access," *Psychological Review*, vol. 105, no. 2, pp. 251–279, 1998.

[11] S. D. Goldinger and T. Azuma, "Episodic memory reflected in printed word naming," *Psychonomic Bulletin & Review*, vol. 11, no. 4, pp. 716–722, 2004.

[12] K. Shockley, L. Sabadini, and C. A. Fowler, "Imitation in shadowing words," *Perception & Psychophysics*, vol. 66, no. 3, pp. 422–429, 2004.

[13] 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.

[14] J. S. Pardo, A. Urmanche, S. Wilman, J. Wiener, N. Mason, K. Francis, and M. Ward, "A comparison of phonetic convergence in conversational interaction and speech shadowing," *Journal of Phonetics*, vol. 69, pp. 1–11, Jul. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0095447017300992

[15] K. Nielsen, "Specificity and abstractness of VOT imitation," *Journal of Phonetics*, vol. 39, no. 2, pp. 132–142, Apr. 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0095447011000027

[16] 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. [Online]. Available: https://asa.scitation.org/doi/10.1121/10.0004373

[17] W. Labov, *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

[18] 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. [Online]. Available: http://link.springer.com/10.3758/BRM.41.4.977

[19] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2022. [Online]. Available: http://www.praat.org/

[20] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, "PsychoPy2: Experiments in behavior made easy," *Behavior Research Methods*, vol. 51, no. 1, pp. 195–203, Feb. 2019. [Online]. Available: https://doi.org/10.3758/s13428-018-01193-y

[21] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced aligner: Trainable text-speech alignment using kaldi," *Interspeech 2017*, 2017.

[22] S. Reddy and J. Stanford, "A web application for automated dialect analysis," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. Denver, Colorado: Association for Computational Linguistics, Jun. 2015, pp. 71–75. [Online]. Available: https://aclanthology.org/N15-3015

[23] J. Kang, "get_vot Praat Script," 2017. [Online]. Available: https://github.com/HaskinsLabs/get_vot

[24] Posit team, *RStudio: Integrated Development Environment for R*, Posit Software, PBC, Boston, MA, 2022. [Online]. Available: http://www.posit.co/

[25] L. Konnelly, "Nuance and normativity in trans linguistic research," *Journal of Language and Sexuality*, vol. 10, no. 1, pp. 71–82, Feb. 2021. [Online]. Available: http://www.jbe-platform.com/content/journals/10.1075/jls.00016.kon

[26] J. Rechsteiner and B. Sneller, "Nonbinary speakers' use of (ING) across gender-related topics," 2021, NWAV 49. [Online]. Available: https://vimeo.com/627644620

[27] L. L. Namy, L. C. Nygaard, and D. Sauerteig, "Gender Differences in Vocal Accommodation:: The Role of Perception," *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 422–432, Dec. 2002. [Online]. Available: http://journals.sagepub.com/doi/10.1177/026192702237958

[28] J. S. Pardo, K. Jordan, R. Mallari, C. Scanlon, and E. Lewandowski, "Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures," *Journal of Memory and Language*, vol. 69, no. 3, pp. 183–195, Oct. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0749596X13000545