WITHIN-SPEAKER PERCEPTION AND PRODUCTION OF DIALECTAL /aɪ/-RAISING

Alyssa Strickler

University of Colorado, Boulder
alyssa.strickler@colorado.edu

ABSTRACT

In the Fort Wayne, Indiana area, speakers are beginning to raise /aɪ/ to /ʌɪ/ preceding voiceless consonants, similar to canonical Canadian raising. Under consideration in this work is the relationship between the perceptual salience of this progressing sound change and speakers' production of the same novel sound. It has been posited that in sound change situations, perception precedes production of novel forms [1, 5]. For the speakers in the current study, however, those who do not yet raise /aɪ/ themselves also do not perceive words with distinctive raising in pre-flap context as consistently as the true raisers. These results indicate that in this situation, individuals in the speech community who are operating within the sound change but not participating in it themselves do not use the raised /aɪ/ at above-chance levels to distinguish otherwise identical words, whereas speakers who *do* raise use the raised /aɪ/ in perception.

Keywords: sound change, perception-production relationship

1. INTRODUCTION

Researchers have studied various forms of /aɪ/-raising, (Canadian raising) in which /aɪ/ is raised to /ʌɪ/ preceding voiceless consonants, throughout the U.S. These raising processes have been described in Philadelphia, Chicago, and in Michigan, among other places [6, 8, 10]. Most previous work with this process has focused on the production of /aɪ/-raising, and only one recent study has experimentally investigated the perceptual salience of the sound as it exists in the northern U.S. [10].

In northeast Indiana, in and around the city of Fort Wayne, /aɪ/ raising is currently believed to be emerging in an incipient form. This transition from a non-raising dialect to one with raising appears to have been caught as it progresses. In the Fort Wayne area dialect, some speakers produce the newer, innovative form (/aɪ/ raised to /ʌɪ/ preceding voiceless consonants), others produce the more conservative, unchanged form of the sound (no /aɪ/-raising in any context), and some produce intermediate forms in which /aɪ/ is raised in certain pre-voiceless phonological environments and not in others, notably not before underlying voiceless flaps. This mix of production patterns within a raising dialect is as of yet relatively unique to the Fort Wayne area, and variation of this type has not been documented as a characteristic of a raising dialect except for the very first description of Canadian raising, in which there was a "Dialect A" and a "Dialect B" found in the speech of school-aged children in Toronto. Speakers of Dialect A produced the word *typewriter* as /tʌɪpɹʌɪɾɚ/, while speakers of Dialect B produced the same word as /tʌɪpɹaɪɾɚ/, without raising in the second syllable [11]. This difference in pronunciation could indicate that while raising may have at some point been phonetically conditioned—Dialect B speakers raised only before the phonetically voiceless /p/ and not before voiced /ɾ/–it eventually became a phonologically conditioned process, exemplified by the Dialect A speakers who also raised before underlyingly voiceless /ɾ/. Notably, by the time that the phenomena was written about again, thirty years later, only Dialect A, the one with phonological raising, was found in Toronto [4].

The most telling phonological environment to observe the perceptual salience of /aɪ/-raising is the pre-flap environment, for two reasons: historically, in Toronto in 1942, it marked a distinction between the production of two groups. It also distinguishes groups of production patterns in the Fort Wayne area. Additionally, other perceptual cues such as vowel duration and final consonant voicing are essentially neutralized. Although there are sometimes differences in vowel length due to underlying voicing of following vowel—i.e., that the /aɪ/ vowel in *writing* is shorter than the vowel in *riding*—vowel length as a perceptual cue in the pre-flap environment in general has been experimentally confirmed to be unreliable [9]. Therefore, it is reasonable to assume that *writing* and *riding* are both pronounced as [ɹaɪɾɪŋ] by speakers with no raising, and are perceptually identical, regardless of possible differences in vowel length. The current study will focus on the pre-flap raisers, and the talkers with no raising.

In perception-focused sound change literature, experimental studies have described that within a single speaker, the relationship between production and perception of an incipient change progresses in one direction—speakers will be sensitive to the

perceptual cues of the novel sound before they start to produce it themselves [1]. This assumption is grounded in claims that in sound change situations, talkers hear innovations around them and then adjust their own pronunciations to resemble the pronunciations of their interlocutors [13]. The current study attempts to determine how an individual's pattern of production of /aɪ/-raising in the pre-flap environment is related to his or her tendency to accurately perceive /aɪ/-raising as the distinguishing feature in otherwise identical words. In other words, is there a consistent relationship between participants' production patterns and which variant(s) of raising they can accurately perceive? Or, are participants able to use an innovative form that they do not produce themselves in identifying these raised and non-raised minimal pair words?

2. METHODS

19 participants (14M, 5F) were recorded in both private homes and semi-public spaces, using a MacBook Pro and an Audio Technica Headset. Participants read a word list 3 times (57 items long, a total of 171 words); this was recorded and analysed using Praat [3]. All words were part of a minimal pair ending in either /t/ or /d/ and were part of a paradigm in which the same two stems could also have -ɾɪŋ ending. For example, one set of words was *bite, bide, biting,* and *biding*. These were chosen as stimuli because there were two sets of minimal pairs, and the bisyllabic words were two-morpheme words rather than monomorphemic. For example, the word *title* was not included in analysis because it is monomorphemic. The /aɪ/ vowels were later hand-annotated in Praat, and time-normalized and measured using FormantPro [16]. Measurements were randomly spot-checked for accuracy. These production data were used to classify each participant as either a pre-voiceless-flap-raiser (i.e., raising in *biting*) or not; the classification of the participant was used when considering the results of the perception task described in the next section.

Classification of participants' raising status was achieved using the production data obtained in the first task. F1 was measured in the third out of ten intervals in the entire diphthong, or at the 30% point of the vowel, a point of measurement that has been used in previous work in this dialect region [2]. Other research in /aɪ/-raising has also used the F1 maximum measure in Hz in the vowel [7]. Close observation of the F1 tracks of the diphthongs in the current study show that the 30% mark of the diphthong is sometimes the F1 maximum, but that this maximum sometimes occurs slightly later, closer to the 40% point in the vowel. Thus, both time points were considered when deciding whether the participant was a raiser or non-raiser. A paired t-test was performed on the first formant measurements at both the 30% point in the vowel and at the 40% point, the paired items being the pre-voiced- and pre-voiceless-flap minimal pairs. There was no case in which the third time point proved significantly different in the statistical tests but the fourth time point did not.

In addition to the production task via wordlist, the same participants completed an identification task in which a single-word stimulus was played through the same AudioTechnica Headset, and participants were presented with two options, the word that was heard and its minimal pair counterpart. Words were displayed on-screen, controlled by PsychoPy script [14]. Participants chose which word they heard by pressing a key. Stimuli included 128 total words, 64 from a talker with no raising at all and 64 from a talker with raising in /aɪ/ in all possible contexts, including when followed by a /t/ flap. Both stimuli speakers were female. Stimuli were naturally produced as part of a previous study with a similar word list reading task and were unedited. The same criteria (paired t-tests between pre-voiced- and pre-voiceless-flap minimal pairs) were used to determine the production patterns of the stimuli talkers.

Figures in the following section were created using ggplot2 [15].

3. RESULTS

3.1. Production Results

To categorize the participants, as described above, data from the word list reading task were used: participants repeated each word three times, and all repetitions were used in analysis, providing that they were valid productions of the intended word. For example, many participants pronounced what was intended to be /baɪdɪŋ/ as /bɪdɪŋ/ or /baɪndɪŋ/, which made their tokens of both *biding* and *biting* unusable, as they could not be compared in paired t-tests and are not included in the average F1 values given below.

A paired, one-tailed t-test was used to determine whether the two sets of productions (underlyingly voiceless /t/-flap and underlyingly voiced /d/ flap) were significantly different within each participant, in order to classify the participant as raiser or non-raiser.

In previous acoustic research, a raised /aɪ/ vowel is defined as a difference of 60 Hertz between the pre-voiceless vowels and the pre-voiced vowels [12, 2]. Using the t-test method described above to determine whether participants produce differences in pre-voiceless formant measures and in pre-voiced formant measures, not all participants in this study produced an average difference of 60 Hz. Participants

were classified as raisers if the results of the paired t-test between their prevoiceless and prevoiced formant measures proved significant ($p<0.05$). Both time point 3 and time point 4 were included in this classification, because the lowest point of the diphthong is sometimes between time points 3 and 4. Where time point 3 was significant while time point 4 was not, that participant was not included with the group of raisers; this happened twice, for S07 and S08. Otherwise, for the other 17 participants, both time point 3 and time point 4 are either both significant or both not. Of the 19 participants, 6 are raisers (5M, 1F) and 13 are non-raisers (9M, 4F).

Table 1: Mean Prevoiceless and Prevoiced F1 measures, t-test results; [R] denotes raiser

Subject	F1 mean prevoiced	F1 mean prevoiceless	t at timepoint3
S01 (M)	640	581	<0.01 [R]
S02 (M)	653	615	<0.01 [R]
S03 (F)	705	680	<0.01 [R]
S04 (M)	601	566	<0.01 [R]
S05 (M)	565	542	0.01 [R]
S06 (M)	674	653	0.02 [R]
S07 (M)	630	611	0.04
S08 (F)	824	763	0.01
S09 (M)	623	603	0.05
S10 (M)	665	659	0.05
S11 (F)	808	783	0.05
S12 (F)	724	697	0.2
S13 (M)	607	638	0.2
S14 (M)	696	702	0.2
S15 (M)	572	579	0.2
S16 (M)	649	644	0.3
S17 (M)	611	603	0.3
S18 (M)	592	591	0.4
S19 (F)	839	838	0.5

2.1. Perception Results

How, then, are production and perception of /aɪ/-raising related within a talker/listener? The following section will discuss how the participants, divided into groups based on their productions from the word list task as described above, performed on perception of the same feature that was measured in their own production.

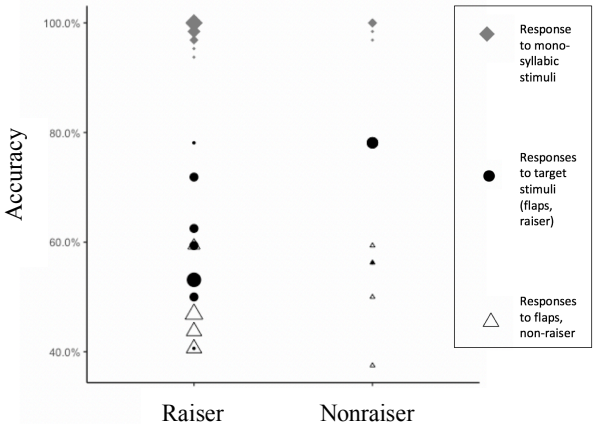
In a word identification task, participants were presented with naturally produced stimuli from one speaker with raising in the relevant pre-voiceless-flap environment, providing minimal pairs contrasting by vowel (/ɹʌɪɾɪŋ/ vs. /ɹaɪɾɪŋ/). In order to address the unlikely possibility that participants may be able to differentiate even non-raised, pre-flap vowels based on vowel length that corresponds to underlying voicing of the following flap phoneme, the vowel lengths of different stimuli from each of the two talkers whose speech was used to provide the stimuli were examined. There is very little difference in average vowel length with respect to underlying voicing of following consonant (difference of 15 ms for the talker without raising and difference of 17 ms for the talker with raising), especially small when compared to the large vowel duration differences in the monosyllabic words due to voicing of the following consonant (difference of 128 ms and 124 ms). So, the previous assertion that participants cannot discern underlying voicing of a flap based on differences in vowel length alone is further supported in these data that were used as stimuli in the identification task [9]. In addition to the minimal differences in vowel length between the pre-voiceless-flap stimuli and the pre-voiced-flap stimuli, all participants performed at chance levels in response to the flapped stimuli spoken by the talker with no raising, as described further below, indicating that vowel length is not a reliable cue to underlying voicing of a following flap.

The stimuli can be described as belonging to one of four categories, based on speaker-type and word-type: there are two stimulus speakers, one raiser and one non-raiser, and two types of words that each speaker produced, monosyllabic words containing /aɪ/ preceding /t/ or /d/, and disyllabic words containing /aɪ/ preceding a flap. The disyllabic words, produced by the talker with raising, were the target stimuli. The stimuli were in four-word paradigms of minimal pairs of monosyllabic words matched with their corresponding disyllabic flapped words (for example, *ride, write, rider, writer*). Response accuracy for all participants was 96-98% correct for the set of monosyllabic stimuli, as expected due to cues of vowel length differences corresponding to following consonant voicing, as well as the presence of voicing cues of the final consonant itself. These stimuli were included in the perception task to ensure that participants were complying with the task, as they should have been very easy to identify. It is assumed that any error in response to the stimuli that were monosyllabic words was not difficulty in discerning the word, but rather a key-press error.

Responses to the disyllable set of stimuli are considered separately, by talker, because the non-raiser stimulus talker produces no difference in these pairs. Before the experiment, binomial distribution tests were run to create a threshold to determine

whether correct responses were above chance or not. This threshold came to be 68% accuracy, indicating that response accuracies below 68% can be attributed to chance. Expectations that responses to the flapped, disyllabic stimuli produced by the talker without raising would be at chance were met: the responses were at or near chance (55% accuracy) because the participants are forced to guess without cues to use to distinguish the underlying voicing of the flap. This reinforces the claim that vowel length is not a useful cue when perceiving vowels in words preceding flaps that are otherwise identical, as in this case, there is no raising to use to distinguish the words either. The talker with raising produces minimal pairs in the disyllabic flapped words as well as in the monosyllabic words, in which the contrasting phoneme is the vowel rather than the voicing of the final consonant. These are the target stimuli for which the responses are reported below. Symbol size reflects number of responses included in the point, where larger symbols represent more participants who performed at this accuracy.

Figure 2: Response accuracy, by stimuli type



Mean accuracy in responses to the target stimuli showed significant differences between the two production groups (raisers at average 73% accuracy, non-raisers at 58% accuracy). The data were modelled using a mixed effects logistic regression model, predicting accuracy of participant responses in the forced-choice word identification task regressed on production group and difference between word frequencies of the options presented, as well as the interaction between the two, and including random effects of participant. In this model, controlling for possible effects of word frequency and the interaction between production group and word frequency, production group was a significant predictor of accuracy. That is, raisers were more likely than non-raisers to respond correctly to stimuli that were identical other than the raised/nonraised vowel, z= 2.086, p=0.03. Differences between options' word frequencies as a predictor in the model was not significant, z=-1.643, p=0.1, indicating that participants did not choose a more frequent word more often than a less frequent word.

A model that included word frequency of the stimulus words as a predictor to accuracy as well as production group did not result in better predictions than the one that only used production group as a predictor ($\chi^2(2)$= 3.331, p= 0.2).

4. CONCLUSIONS

This study investigated the within-speaker relationship between production and perception of an ongoing sound change in Fort Wayne, Indiana, where speakers are beginning to raise /aɪ/ to /ʌɪ/ preceding voiceless consonants. Specifically, in this paper, the relevant phonological context of this raising is in the underlyingly-voiceless-flap environment (ɹʌɪɾɚ), because it is a case in which some members of the speech community produce the novel form and raise in the word *writer* and others do not. In this group of participants, it happens that 6 of 19 are pre-voiceless-flap raisers; it is unknown how this compares to the distribution of non-pre-voiceless flap raisers in Toronto in 1942 [11].

Here, those that raise in the pre-voiceless-flap phonological context were here defined as raisers, and those who do not were called non-raisers. The raisers (6) performed significantly better than the non-raisers (13) in the perception task, in which they identified target stimuli that were produced by a raiser. The target stimuli isolated the raising cue from other possible cues, to ensure that participants were in fact responding to the raising, and not using other acoustic cues to identify the words. These results indicate that in this instance, the participants from the speech community in which the sound change is occurring who produce the novel version themselves *do* perceive it at higher rates of accuracy than the participants from the same speech community who do not produce it, though they are exposed to it.

Thus, the intra-speaker relationship between production of /aɪ/-raising and perception of /aɪ/-raising seems to be that speakers in the Fort Wayne area do, at this point, only use the form that they produce in perception of /aɪ/. Although production results strongly suggest that this is a sound change in progress, as there is some evidence that this type of raising is more common among younger speakers in the Fort Wayne area, this is an aspect of the work that requires further development [2]. It does not seem that speakers are able to perceive a more advanced form than they produce, though more work is necessary to fully support this claim.

5. REFERENCES

[1] Beddor, P. S. 2015. The relation between language users' perception and production repertoires. Proceedings of the 18th ICPhS, 171-204.

[2] Berkson, K., Davis, S., Strickler, A. 2017. What does incipient/ay/-raising look like?: A response to Josef Fruehwald. *Language*, *93*(3), e181-e191.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer software]. Version 6.0. 37.

[4] Chambers, J. K. 1973. Canadian Raising. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 18(2), 113-135.

[5] Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., Wissing, D. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *J. Phon.*, *66*, 185-216.

[6] Dailey-O'Cain, J. 1997. Canadian raising in a midwestern US city. *Language Variation and Change*, *9*(1), 107-120.

[7] Fruehwald, J. 2016. The early influence of phonology on a phonetic change. *Language*, 92(2), 376-410.

[8] Fruehwald, J. 2013. The phonological influence on phonetic change. University of Pennsylvania.

[9] Herd, W., Jongman, A., Sereno, J. 2010. An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *J. Phon.*, *38*(4), 504-516.

[10] Hualde, J. I., Luckina, T., Eager, C. D. 2017. Canadian Raising in Chicagoland: The production and perception of a marginal contrast. *J. Phon.*, 65, 15-44.

[11] Joos, M. 1942. A phonological dilemma in Canadian English. *Language*, 18, 141-144.

[12] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*, 216-226.

[13] Ohala, J. J. 1981. Articulatory constraints on the cognitive representation of speech. In *Advances in Psychology* (Vol. 7, pp. 111-122). North-Holland.

[14] Peirce, J.W. 2007 PsychoPy - Psychophysics software in Python. *J. Neurosci. Methods*, 162(1-2):8-13.

[15] Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

[16] Xu, Y. 2015. FormantPro.praat. Available from http://www.phon.ucl.ac.uk/home/yi/FormantPro/.