Laboratory Phonology
Journal of the Association for Laboratory Phonology

Hashimoto, D., & Hay, J. (2026). A system-wide effect in cross-dialect convergence via social category activation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 17(1), pp. 1–46. DOI: https://doi.org/10.16995/labphon.18216

Open Library of Humanities

A system-wide effect in cross-dialect convergence via social category activation

Daiki Hashimoto*, Joetsu University of Education, Yamayashiki 1, Joetsu, Niigata, Japan, daiki@juen.ac.jp

Jennifer Hay, University of Canterbury; New Zealand Institute of Language, Brain and Behaviour

*Corresponding author.

This study explores the question of how phoneme-specific shadowing effects are. When shadowing a vowel from another known dialect, are the effects limited to the specific vowel to which the speaker is exposed? Or does more general knowledge of the shadowed dialect have an effect on the wider vowel space? Twenty-four New Zealand English (NZE) speakers participated in a speech elicitation experiment. In the shadowing session, they repeated model speech stimuli including DRESS vowels that were produced by an American English (AmE) model talker. In the preshadowing and postshadowing sessions, they read aloud DRESS words and TRAP words. It was found that the NZE speakers increased F1 values of DRESS in the shadowing session as compared with the preshadowing session; that is, they shifted their vowel formant values towards those of the AmE model talker. Likewise, they increased the F1 values of TRAP in the postshadowing session, despite the fact that they did not repeat the model stimuli with TRAP in the shadowing session. This suggests that the imitation of AmE DRESS led to general social category activation of "American English", which has an impact on the postexposure production of TRAP.

Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by the Open Library of Humanities. © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/. Articles published by the Open Library of Humanities may also contain other CC-licensed material, or material used or reproduced under an exception or limitation to copyright. Such material is not subject to the article's license. It is either marked with its own license, or marked that it has been reproduced with permission or under fair-use.

OPEN ACCESS

1. Introduction

It is well known that if speakers are exposed to a variant from another dialect, they may converge on their pronunciation of that variant (see Section 1.1). Less well known is how such exposure might affect the wider linguistic system, as opposed to the specific variant to which they are exposed. We investigate this question in two ways, by exploiting a cross-dialectal overlap in the phonetic realization between two different vowels. The particular variation of concern here is the vowel phoneme in the word *dress*, which is produced as a mid vowel [ɛ] in American English (AmE), whereas it is produced as a high vowel [i] in New Zealand English (NZE). Likewise, the vowel phoneme in the word *trap* is produced as a low vowel [æ] in AmE, whereas it is produced as a mid vowel [ɛ] in NZE. NZE is known to have undergone a "short front vowel chain-shift" (Hay et al., 2015), and TRAP and DRESS vowels shifted upward in the history of NZE (Gordon et al., 2004; Bauer & Warren, 2008). The extent of the vowel differences is illustrated by Hay et al. (2008: 24), who note that "If a New Zealander says they have a *pet cat* it can sound like a *pit ket*, or even a *peat ket*," and Hay et al. (2015) note that "NZE *bat* often sounds to speakers from elsewhere like *bet*, NZE *bet* sounds like *bit*." Ladefoged & Johnson (2015: 234) note that "in some forms of New Zealand English, the vowel in *sacks* is similar to the American English [ɛ] vowel in *sex*." **Figure 1** illustrates the cross-dialect differences in DRESS and TRAP vowels.

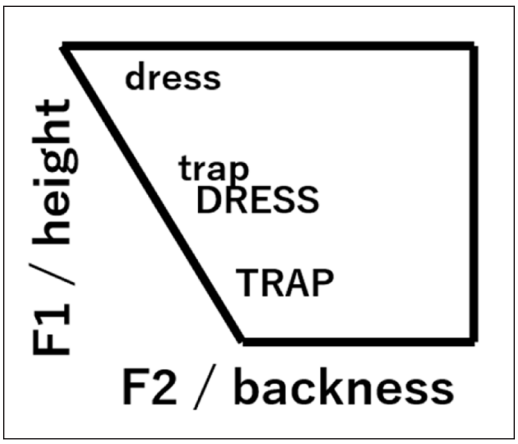

Figure 1: Schematic distribution of DRESS and TRAP vowels in AmE (capital letters) and NZE (lowercase letters).

How, then, will New Zealander's shadowing of an AmE DRESS vowel affect, and be affected by, the TRAP vowel? In a cross-dialectal shadowing task, we ask (1) whether the pronunciation of TRAP is affected, after shadowing DRESS, and (2) whether minimal pairs with TRAP can affect the production of a shadowed DRESS vowel. At the core of both of these questions is the question of how specific the effects are when shadowing a vowel from another known dialect. Are the effects limited to the specific vowel to which the speaker is exposed? Or does more

general knowledge of the shadowed dialect have an effect on the wider vowel space? The current study provides the evidence that a system-wide effect is caused by social category activation. Before discussing these questions in more depth, we review some of the core literature on phonetic convergence.

1.1. Phonetic convergence

Convergence is a phenomenon in which an individual shifts their behaviour towards another individual or group. It is observed in many behavioural domains (Mercado III et al., 2014), including speech. Convergence in pronunciation is sometimes referred to as phonetic convergence. A speaker is likely to be influenced by the phonetic variants used by surrounding speakers. Evans & Iverson (2007) provide a good case study of the long-term effects of phonetic convergence. They explored how Northern England English accent speakers modify their speech when in contact with Standard Southern British English speakers. It is well known that the FOOT-STRUT split and TRAP-BATH split did not occur historically in Northern England English (Wells, 1982: 196; 232); i.e., the word *cud* is homophonous with the word *could* in Northern England English, and the word *bath* is produced with the same vowel phoneme as the word *trap*. Evans & Iverson demonstrated that Northern accent speakers gradually changed their formant values of the STRUT and BATH vowels after they attended universities, where they were likely to interact with Standard Southern British English speakers. Their habitual formant values slightly shifted towards those of the STRUT and BATH vowels in Standard Southern British English.

Such long-term convergence arises as the cumulative effect of online convergence effects across multiple utterances. A speaker can adjust the probability of producing a particular phonetic variant in accordance with an interlocutor. For example, a Norwich speaker is known to stylize their speech by choosing one of the socially meaningful variants such as [t] or [ʔ] in the production of /t/ in word-medial position (e.g., *but̲t̲er* and *wa*t̲*er*). Bell (1984) shows that a speaker changes the rate of /t/ glottalization depending on how often the interlocutor produces the glottal variant [ʔ] during the conversation. To take another example, a French speaker chooses one of the two phonetic variants [dʒ] and [ʒ] in the production of an affricate-initial loanword such as *jeans*. Lev-Ari & Peperkamp (2014) demonstrate that a French speaker changes the rate of selecting the phonetic variant [dʒ] depending on how often their interlocutors produce the variant during a spontaneous language game. There is abundant evidence that a speaker converges towards an interlocutor during interactive tasks such as map tasks (Pardo, 2006; Cao, 2023) and diapix tasks (Kim et al., 2011; Biro et al., 2022).

Phonetic convergence is also observed even when the interlocutor is not physically present. Many experimental studies demonstrate that a speaker adjusts their pronunciation towards a model talker, when and after the speaker shadows (repeats) model speech stimuli (Goldinger, 1998; Shockley et al., 2004; Babel, 2009; Fiasson, 2015; Pardo et al., 2017; Kwon, 2019; MacLeod

& Di Lonardo Burr, 2022), and after they listen to model speech stimuli (Goldinger, 2000; Goldinger & Azuma, 2004; Delvaux & Soquet, 2007; Nielsen, 2011; Yu et al., 2013; Hashimoto et al., 2022). In these experimental studies, the participants first read aloud target words, and the recordings are treated as baseline utterances. Following this, the participants are exposed to model speech stimuli through a shadowing task or a listening task. Some studies analyse shadowed utterances in order to assess convergence. On the other hand, some studies employ a postexposure session, where the participants read aloud target words as in the baseline session, and the utterances in the postshadowing session are the target of analysis. Comparing the baseline utterances versus shadowed utterances (or postexposure utterances) allows researchers to explore lab-based speech convergence. Previous studies assess the similarities between model stimuli and shadower's utterances using either a perceptual experiment (i.e., an AXB experimental paradigm) or an acoustic analysis of a particular phonetic feature (e.g., formant values, spectral peak values, VOT, segmental duration, f0, and intensity). The studies cited above all demonstrate that shadowed utterances or postexposure utterances are more similar to model speech stimuli than baseline utterances are. These results can be interpreted to mean that speakers converge towards variants produced by model talkers once they are exposed to the model speech stimuli of these variants in non-interactive tasks.

These findings suggest that a speaker's mental representations can be continuously restructured through imitative tasks (Goldinger, 1998; Tilsen 2009). Exemplar Theory provides one straightforward account for lab-based phonetic convergence. Upon exposure to model speech stimuli, unique traces may be stored in the listener's memory, and the episodic memories or "exemplars" retain phonetically rich information (Johnson, 2006). In speech production, multiple exemplars are chosen as "contributors" and averaged together to form a production target. Exemplars with higher resting activation levels are more likely to be chosen as contributors than those with lower resting activation levels (Pierrehumbert, 2001; Wedel, 2006). The newly stored exemplars have relatively high resting activation levels due to the recency of storage. Hence, they are likely to contribute to the creation of a production target afterwards, and therefore the shadowed and postshadowing utterances are likely to have phonetic features that are similar to the model speech stimuli.

In many of the previous studies about lab-based convergence, the model talker(s) and participants in the experiments share the same language background; i.e., they speak the same language and dialect. Some previous studies explore "cross-dialect convergence," in which model talkers and participants speak different dialects of the same language (Delvaux & Soquet, 2007; Babel, 2010; Dufour & Nguyen, 2013; German et al., 2013; Walker & Campbell-Kibler, 2015; Gnevsheva et al., 2021; Clopper et al., 2024; Lee-Kim & Chou, 2024). These previous studies have demonstrated that a speaker shifts their production towards a model talker with a different dialectal background when and after shadowing model speech stimuli. Cross-dialect convergence

offers an excellent opportunity to develop our understanding of how novel variants are treated by the cognitive system and how they are mentally stored in relation to sociolinguistic categories. The overarching goal of this study is to investigate the mental representations and cognitive mechanisms that are involved in phonetic convergence.

Based on the above convergence literature, we can safely assume that such a cross-dialectal convergence effect will be observed in New Zealanders. As noted at the outset of this study, the DRESS vowel is produced as a high vowel [i] in NZE, whereas it is realized as a mid vowel [ɛ] in AmE. Accordingly, when NZE speakers are asked to shadow AmE DRESS words, they should produce a lower vowel (with higher F1 values) than in their baseline production. Such convergence effects may also persist in postshadowing productions. These input-driven convergence effects are predicted by hypotheses concerning the continuous restructuring of mental representations reviewed above. However, an important open question is whether the effect is limited to the DRESS vowel, or whether shadowing DRESS also affects other vowels. In the next section we review literature suggesting that social knowledge can lead to more general system-wide priming effects.

1.2. Imitation of a particular vowel phoneme from a different dialect may be generalized to another vowel phoneme in postexposure production

Previous literature suggests that exposure to a social concept (or a different dialect) shifts the perception and production of a phoneme. For example, Hay et al. (2006) demonstrate that speakers change their perception of KIT vowels upon seeing the word associated with a dialect. The participants in their experiment listened to continua of the target vowels (KIT, DRESS, and TRAP). The formant values were acoustically manipulated, and they ranged from Australian-English-like variants to NZE-like variants. Their task was to indicate which token in the continuum best-matched the target vowel in the word underlined in their answer-sheet. The word “New Zealander” was written at the top of each answer sheet in one condition, and the word “Australian” was written in the other. They found that the female participants were more likely to choose an NZE-like KIT vowel token when “New Zealander” was written on the answer sheet than when “Australian” was written on it, even though most of the participants believed that the talker was a New Zealander. These results suggest that seeing a regional label associated with a particular dialect influences vowel perception.

Several studies show that activation of a dialect region can affect production. Love & Walker (2013) and Sanchez et al. (2015) explored spontaneous speech, and found that speakers shift their production towards a particular dialect when discussing topics associated with the dialect. More specifically, Love & Walker (2013) demonstrate that English speakers living in the United States (US) produce more AmE-like rhotic variants when talking about American football than when talking about English soccer, and Sanchez et al. (2015) report that NZE speakers produce KIT

vowels with Australian-English-like F1 values when talking about Australia than when talking about something neutral. The results of these two studies indicate that talking about something related to a particular dialect leads to the shift towards the dialect.

Wade (2022) provides evidence that AmE speakers shift their speech production in accordance with their expectations regarding linguistic behaviours of a talker. The participants in their research read aloud target words including PRICE /aɪ/ by guessing the target words based on their definitions. The clues were visually displayed on the PC screen in the baseline session, whereas they were auditorily provided in the exposure session. The talker was either a Southern talker or a Midland talker in the exposure session. The participants never heard PRICE vowels in the exposure session. The results of their experiment demonstrate that PRICE is more likely to be produced with Southern-like, glide-weakened tokens, in the exposure session than in the baseline session, if the clues were provided by the Southern talker in the exposure session. These results suggest that exposure to Southern-like features except PRICE vowels influences the phonetic realization of PRICE vowels, providing evidence that activating a dialect which is familiar to the participant can have system-wide effects.

Results like those above, and similar, have led multiple authors to posit that social information and phonetic information are stored together, and are linked (see. e.g., Sumner et al. 2014). An intuitive account of how this could work can be captured by a mechanism we will refer to here as "social category activation." English speakers may have opportunities to hear a variety of dialects, for example through social media and TV. Hence, the mental lexicon may store exemplars that encode speech from other dialects. Exemplars form an exemplar cloud on the basis of sociolinguistic characteristics (Foulkes & Docherty, 2006), and they may be linked with social categories such as AMERICAN and AUSTRALIAN in the lexicon. The exposure to a social concept raises the activation of the relevant social category, in turn increasing the activation levels of the exemplars associated with that social concept (Drager & Kirtley, 2016). Consequently, the exemplars associated with the social category become more readily available afterwards; i.e., the speech perception and production shift towards the variants associated with that social concept to some extent. Such an effect has not been tested in the context of a shadowing paradigm, but the above results suggest that it should be observed. The current study therefore addresses the following key question in order to test such an effect in a shadowing paradigm.

(1) Key research question 1
Will a cross-dialectal shadowing task lead a participant to adjust their pronunciation of a vowel that they have not been exposed to?

It may be possible that imitation of a particular vowel from a different dialect generalizes to another vowel phoneme. Nielsen (2011) demonstrated that the imitation of an acoustically manipulated token can be generalized to words and phonemes that participants did not listen to. She ran an experiment in which participants listened to words beginning with [p^h] that were

acoustically modified to have extremely long VOT values. The participants read aloud not only the words with word-initial /p/ but also other words with word-initial /p/ or /k/ in the preexposure and postexposure sessions. She compared the preexposure tokens with the postexposure tokens, and found that the unheard words were produced with longer VOT values, along with the heard words, in the postexposure session than in the preexposure session. In addition, the results show that the participants produced /k/ as well as /p/ with longer VOT values in the postexposure session than in the preexposure session. These results provide the evidence that the imitation can be systematically generalized at abstract levels to the words and phonemes that participants did not listen to during the experiment (see also Fiasson, 2015; Kwon, 2019). The current study aims to extend these previous findings to cross-dialect convergence. More specifically, we investigate whether imitation of a particular vowel indexed to a different dialect is systematically generalized to another vowel phoneme that is likewise indexed to that dialect.

The generalization of imitation from one vowel phoneme to another is predicted by the hypotheses concerning the mental mechanism of social category activation reviewed above. Under this account, exemplars are assumed to be associated with social categories in memory, and exposure to a social concept may trigger spreading activation between social categories and their associated exemplars. This mechanism makes a prediction that a shadower may modify the phonetic values of a vowel phoneme to which the shadower was not exposed in the shadowing session. This is because exposure to a particular phoneme from a different dialect in a shadowing session may raise the activation of the relevant social category, and thereby exemplars associated with the social category may become more readily available in the postshadowing session. Consequently, the postexposure production may shift towards the different dialect in general. Testing this prediction will contribute another piece of evidence to the growing body of literature documenting the dialect priming effects on speech, and explicitly test the degree to which there are effects in shadowing beyond the particular shadowed variant.

As noted at the outset of this study, NZE provides an interesting test case to address the key question. Many young NZE speakers are exposed to and familiar with AmE to some extent even if they have never lived in the US. This observation has been made by some previous studies. For example, Walker & Campbell-Kibler (2015) note that "New Zealanders have a large amount of exposure to American accents through popular media," and they demonstrated that NZE speakers could identify a talker as an AmE speaker when they listened to AmE accents. Gibson (2019, 2024) notes that the use of AmE accent is default for NZ singers when they play pop music and Gibson & Hay (2025) show that music specifically leads NZ listeners to have a lower DRESS-TRAP boundary. These observations suggest that NZE speakers may store exemplar memories that encode AmE speech to some extent, and the exposure to AmE model speech stimuli may thus raise the activation of the social category "American English." As New Zealanders are familiar with AmE, then exposure to one vowel in a shadowing task would be predicted to have potential effects on the later production of other vowels that differ between the varieties.

In particular, the current study explores whether NZE speakers modify the phonetic values of the TRAP vowels after shadowing DRESS vowels. As discussed above, the TRAP vowels in NZE are generally higher than their correspondences in AmE. Using this test case, the general prediction presented above can be narrowed down. Our specific prediction is that NZE speakers may produce TRAP vowels with a lower tongue position (with higher F1 values) after shadowing the model speech stimuli including DRESS vowels produced by an AmE model talker. In the experiment reported below, NZE speakers were asked to shadow words with DRESS vowels that were produced by an AmE model talker, and never heard TRAP words from the model talker. They were asked to produce TRAP words in the preexposure and postexposure sessions.

1.3. Presence of a minimal pair competitor may affect the extent of imitation

It is well known that words with higher neighbourhood density (i.e., more minimal pair competitors) are more difficult to recognize and process in speech perception. Luce & Pisoni (1998) is a classic study that explored minimal pair effects on word recognition. They ran an auditory lexical decision task (Experiment 2). Half of the words employed in their experiment were real words, and the other half were nonce words. The words varied in terms of neighbourhood density. The participants listened to audio stimuli, and decided whether the heard word was real as quickly as possible. Their results demonstrated that real words with many neighbours were recognized more slowly than words with few neighbours. Vitevitch & Luce (1998) ran a word repetition task to study the effects of minimal pair competitors on word recognition. The participants in their experiment were asked to repeat a word as quickly as they heard the word. Their results showed that words with a larger number of minimal pair competitors were repeated less quickly.

In speech production, words with minimal pair competitors are more likely to be hyperarticulated than those without minimal pair competitors. Baese-Berk & Goldrick (2009) ran a speech elicitation experiment (Experiment 2), in which participants read aloud target words beginning with voiceless stop consonants /p, t, k/. Half of the target words had a minimal competitor word with a voiced stop consonant, but the others did not. Their task was to instruct a listener to “Click on the (target word)” while seeing two other dummy words on the PC screen. Sometimes, one of the dummy words was a minimal pair competitor word beginning with a voiced stop. Their experimental results show that speakers produced word-initial /p, t, k/ with longer VOT values in relation to target words that have minimal competitor words in terms of word-initial voicing than those which do not have minimal pair competitor words. In addition, they demonstrated that speakers produced word-initial /p, t, k/ with longer VOT values when seeing a minimal pair competitor word as a dummy word than without seeing a minimal pair competitor word. Stern & Shaw (2023) conducted a replication of this finding. They employed pseudowords instead of real words as target words. The pseudowords included word-initial voiceless stop

consonants, some of which had lexical minimal pair competitors in terms of word-initial stop consonant voicing. Their task was to instruct a listener to "type the (target word) number" while seeing a target word along with dummy words. Dummy words included a minimal competitor word beginning with a voiced stop consonant in some trials. Their results show that the VOT values of the word-initial stop consonants were enhanced by a minimal pair competitor word on the screen only when the competitor was a real word.

These minimal pair effects also occur in spontaneous speech production. Wedel et al. (2018) explored natural speech data, and analysed the phonetic realizations of stop consonants /p, t, k, b, d, g/ and six nonback vowels /i, ɪ, e, ɛ, ʌ, æ/. They found that the voiceless stop consonants /p, t, k/ are produced with longer VOT values if there is a minimal pair competitor in initial-stop voicing in the English lexicon. Likewise, the voiced stop consonants /b, d, g/ are produced with shorter VOT values if there is a voiceless stop minimal pair competitor. Furthermore, their statistical results suggest that the six nonback vowels are produced with more contrastive formant values if there is a minimal pair competitor in at least one of the neighbouring vowels.

There have been a variety of hypotheses to account for the fact that words with minimal pair competitors are hyperarticulated in speech production. One of the most straightforward accounts is that a speaker has a bias ensuring the intelligibility of a speech signal, and makes an effort to maintain the distinctiveness of linguistic units such as words and phonemes. A speaker is hypothesized to increase the intelligibility of a speech signal when they think that the message transmission is unlikely to be successful (Lindblom, 1990; Pate & Goldwater, 2015; Hall et al., 2018). This bias is referred to as the *intelligibility bias* in what follows. As noted in the beginning of this section, words with minimal pair competitors tend to be more confusing than those without minimal pair competitors. Hence, a speaker may hyperarticulate words to differentiate them from minimal pair competitor words, and thereby increase the accuracy of message transmission.

Another account comes from the memory storage process. Previous literature argues that phonetically ambiguous words are occasionally discarded and not stored as new exemplars in memory (Wedel, 2006; Sóskuthy, 2015; Todd et al., 2019). If words with minimal pair competitors are produced with reduced speech signals, which are thereby similar to their competitors, they may be underrepresented in the memory storage process. Consequently, clear speech signals that are distinct from minimal pair competitor words tend to leave unique traces in memory. On the other hand, this kind of confusion does not take place in relation to words without minimal pair competitor words. In consequence, speech signals with various degrees of intelligibility can be stored in memory with regard to words without minimal pair competitor words. Through this memory storage process, lexical categories that have minimal pair competitor words are represented with a larger number of exemplars that encode hyperarticulated forms than those which do not have minimal pair competitor words. We note that these two accounts (i.e., the

intelligibility bias account and the memory storage account) are not mutually exclusive, and make a similar prediction regarding the minimal pair effects on hyperarticulation.

As noted in Section 1, one goal of this study is to explore whether minimal pair effects can be extended to cross-dialect phonetic convergence. Dialects may differ in the phonetic realization of a given phoneme (Wells 1982: 73); as a result, certain words said in a particular dialect sound similar to different words in another dialect. If imitating model speech stimuli that are produced by a talker with a different dialect background has a risk of blurring a lexical distinction between minimal pair words, do shadowers nonetheless imitate the model speech stimuli to the same degree? To the best of our knowledge, the effects of minimal pairs have not yet been examined in the context of cross-dialect convergence.

(2) Key research question 2
Does the presence of a minimal pair affect the degree of shadowing in a cross-dialectal shadowing task?

NZE provides an interesting test case to address the key question. As reviewed in the outset of this study, AmE DRESS vowels are perceptually close to NZE TRAP vowels. This proximity suggests that AmE DRESS vowels may be ambiguous between the two vowel phonemes, potentially making it difficult to identify the intended words. The current study explores the extent to which NZE speakers converge towards AmE model talkers in relation to DRESS depending on the existence of a minimal pair competitor word with TRAP. If NZE speakers converge towards AmE speakers in their pronunciation of DRESS vowels, the contrast between DRESS and TRAP vowels may become less clear than usual. This situation may impair the lexical intelligibility of some minimal pair words between the two vowels (e.g., *bed* vs. *bad*) to some extent. On the other hand, the imitation of AmE DRESS vowels may not affect the lexical intelligibility of certain words, such as *deaf* and *jet*, as they do not have minimal pair competitors with TRAP vowels such as **daf* and **jat*.

Importantly, our prediction with respect to this question is conditional on the results of question (1).[1] If shadowing AmE DRESS has no effect on the production of the New Zealander's TRAP vowel, this suggests that there is not a system-wide effect caused by social category activation. If AmE DRESS vowels and NZE TRAP vowels are simultaneously active, then the AmE DRESS vowels may be ambiguous between the two vowel phonemes, potentially making it difficult to identify the intended words. Lexical ambiguity arises particularly for DRESS words

[1] What was specifically preregistered was: "New Zealand English (NZE) speakers may produce DRESS vowels with a lower and backer tongue position (with higher F1 and lower F2 values) when and after shadowing the model speech stimuli including DRESS vowels produced by an American English (AmE) model talker. This imitation effect may be inhibited by the presence of a minimal pair competitor with TRAP vowels." At the time of the preregistration, we had not properly registered that the degree to which predictions for this effect could be dependent on the answer to research question 1, and we expected to observe a minimal pair effect.

that have minimal pair competitors containing TRAP vowels. The imitation of AmE DRESS may affect the lexical intelligibility of some DRESS words (e.g., *bed* and *guess*) which have minimal competitor words with TRAP (e.g., *bad* and *gas*), whereas it may not affect the lexical intelligibility of some DRESS words which do not have minimal competitor words with TRAP (e.g., *deaf* and *jet*). Based on the hypotheses concerning the memory storage process and the intelligibility bias reviewed above, our prediction in this case is that NZE speakers may converge to AmE DRESS vowels in the words that have minimal pair competitor words with TRAP vowels to a lesser extent than in the words that do not have minimal pair competitor words with TRAP vowels. More specifically, NZE speakers may produce DRESS vowels with a higher tongue position (with lower F1 values) in relation to DRESS words that have minimal pair competitor words with TRAP vowels than DRESS words without minimal pair competitor words with TRAP vowels, when and after shadowing AmE model speech stimuli with DRESS.

However, if speakers shift their TRAP vowel in the postshadowing session, then this suggests that exposure to AmE English has activated both AmE DRESS and TRAP vowels. In this case, the speakers have shifted to an overall more AmE-mode, and thus the contrast between the two vowels would be maintained and not in danger. Under this scenario, DRESS imitation would occur independently of the presence of a minimal pair competitor containing TRAP. Specifically, no difference in F1 values would be expected between DRESS words with and without TRAP minimal pair competitors in either the shadowing or postshadowing sessions.

In sum, our predictions for the two key research questions are interlinked. The degree to which this effect is predicted is deeply entangled with the question of whether exposure to the AmE DRESS vowel just affects the speaker's production of DRESS, or whether it shifts the whole vowel space they are operating with in the direction of AmE vowels. If it is just the DRESS vowel, then the local-variety minimal pair TRAP words are highly overlapping with the target DRESS, and we would expect the intelligibility bias and memory storage process to impede the degree of convergence. But if the whole vowel space is affected and AmE TRAP vowels are thus also activated, then the integrity of the lexical contrast remains intact, and the produced DRESS vowel is not in danger of overlapping with the activated AmE-like TRAP representations. As New Zealanders have reasonable experience with AmE, and the literature shows evidence from other tasks that social category activation can have system-wide effects (see Section 1.2), we regard the latter to be the most likely.

2. Methodology

In order to address the research questions, we ran a shadowing experiment. The experiment consists of five key sessions: baseline session, shadowing session I, postshadowing session I, shadowing session II, and postshadowing session II. The participants read aloud target words with DRESS vowels and TRAP vowels in the baseline and postshadowing sessions, and shadowed

words with DRESS vowels that were produced by an AmE model talker (target condition) and a NZE model talker (control condition) in the shadowing sessions. That is, the participants never heard the TRAP vowels from either model talker in the experiment. The order of the target and control conditions was counter-balanced between participants.

Along with the predictions provided in the preceding section, the methodology illustrated in this section was preregistered using AsPredicted (https://aspredicted.org/index.php) prior to beginning the experiment. The ID is #162355. However, several changes were made to the analyses following data inspection (see Footnotes 1, 3, 4, and 5). *Laboratory Phonology* did not review this study prior to data collection.

2.1. Participants

Twenty-four female NZE speakers participated in the current experiment as shadower participants. They identified themselves as NZE speakers. The age of the participants ranged from 18 to 34. We limited our data collection to female speakers to control for any gender effects. It is known that gender plays some role in vocal imitation (Namy et al., 2002; Babel, 2009; Babel et al., 2014; Lee-Kim & Chou, 2024), but our interest does not lie in the gender effects.

2.2. Stimuli

Monosyllabic English words including DRESS and TRAP vowels were first collected as target word candidates using CELEX2 (Baayen et al., 1995). Amongst them, 80 words were chosen as target words. There were 40 words containing each of the target vowels DRESS and TRAP that have a minimal pair competitor with the other target vowel, and 40 words including DRESS or TRAP that do not have a minimal pair competitor with the other target vowel. The existence of a minimal pair competitor was identified using CELEX2. There were three exclusion criteria. First, we excluded words with prenasal vowels. This is because prenasal TRAP vowels are known to be produced with a higher tongue position in AmE (Labov et al., 2006: 84). Second, we discarded words with vowels followed by /l/ and /ɹ/. The reason is that vowel-to-approximant boundaries are not always clear, and allophonic variation is triggered by the following /l/ and /ɹ/ (Di Paolo et al., 2010: 88). NZE and AmE can also differ in the degree to which these phonemes are present, as NZE is largely nonrhotic, and can have extensive /l/-vocalization (Hay et al., 2008: 35). We removed words with extremely high word frequency such as *have* and *that*. This is because word frequency is known to influence the degree of imitation (Goldinger, 1998; Babel, 2010; Pardo et al., 2017; Hashimoto et al., 2025). The list of target words is provided in **Table 1**.[2]

[2] An anonymous reviewer questioned whether *pest* and *vest* may have minimal pairs with TRAP vowels, such as *past* and *vast*. Note that the words *past* and *vast* belong to the BATH lexical set rather than TRAP, and they contain /ɑː/ rather than /æ/ in NZE.

DRESS with minimal pair (20 words)	beck, bed, beg, bet, dead, fed, flesh, guess, heck, leg, mesh, mess, met, neck, peck, pet, set, sec, tech, vet
DRESS without minimal pair (20 words)	check, chef, chess, chest, deaf, debt, deck, desk, fetch, guest, hedge, jet, next, pest, sketch, speck, step, test, text, vest
TRAP with minimal pair (20 words)	back, bad, bag, bat, dad, fad, flash, gas, hack, lag, mash, mass, mat, knack, pack, pat, sat, sack, tack, vat
TRAP without minimal pair (20 words)	cat, chat, dash, fact, fat, hat, hatch, jab, jazz, mad, map, match, nap, pad, smack, snack, snap, stack, tap, tax

Table 1: List of target words.

The shadower participants were asked to read aloud the 80 target words in the baseline and postshadowing sessions, and shadow the 40 target words with DRESS vowels in the shadowing sessions (see Section 3.3 for the detail). In the baseline and postshadowing sessions, they were also asked to read aloud 10 filler words (*butter, city, dirty, eighty, fighter, party, pretty, shutter, tutor,* and *water*). The filler words largely differ from the target words in terms of the phonological structure. They include neither DRESS nor TRAP vowels, and are disyllabic words. In the shadowing sessions, 16 filler words were added to the word list. Eight of them included LOT vowels (*box, fox, pot, shop, shot, socks, stop, top*), and the others included FLEECE vowels (*beach, cheese, deep, seed, seek, speed, speak, teeth*). These words were employed as filler words because shadowing these words may allow a shadower participant to capture the vowel space of a model talker, and distract them from the target vowels. The two vowels are two of the most peripheral vowels in the two varieties of English (i.e., FLEECE is the highest front vowel, and LOT is the lowest back vowel).

The 56 shadowing words were presented to two model talkers. One talker comes from California, the US, and the other comes from New Zealand. They were both young female speakers. They were individually recorded reading aloud the 56 words at a sound-attenuated room in Joetsu University of Education in Japan, equipped with a recorder (TASCAM DR-40X) and a head-mounted microphone (AKG C-520). The sampling rate was set to 44,100 Hz, and the resolution was 16 bits. The recordings of the 56 shadowing words were used as model speech stimuli in the shadowing experiment. The intensity of the sound files was equalized using a Praat script.

Along with the 56 shadowing words, they were also asked to read aloud a variety of words with multiple vowels. Based on the recordings, **Figure 2** was generated. The left-hand figure shows the vowel space of the AmE model talker, and the right-hand figure shows that of the NZE model talker. The vowel space was defined on the basis of only nonrhotic monophthongs, and the formant values were converted into Bark (Traunmüller, 1990). See Section 2.4 for the formant measurement and Bark scaling. Notice that AmE and NZE are different in terms of their

phonemic systems; i.e., AmE has nine non-rhotic monophthongs (KIT, DRESS, TRAP, STRUT, FOOT, FLEECE, LOT, THOUGHT, and GOOSE) while NZE has eleven (KIT, DRESS, TRAP, STRUT, LOT, FOOT, FLEECE, NURSE, START, THOUGHT, and GOOSE). Note that the accent of the AmE talker may have a vowel system with eight monophthongs due to the cot-caught merger (Wells, 1982: 473–475; Labov et al., 2006: 58; Ladefoged & Johnson, 2015: 97). As is clear in the figures, the location of the NZE model talker's DRESS vowel is relatively higher than that of the AmE model talker's DRESS vowel, and that of the NZE model talker's TRAP vowel is relatively higher than that of the AmE model talker's TRAP vowel. In addition, the acoustic values of AmE DRESS (Bark-scaled F1: 7.6, Bark-scaled F2: 12.7) and those of NZE TRAP (Bark-scaled F1: 6.9, Bark-scaled F2: 13.6) are relatively close to each other. These vowel distributions are in line with the descriptive generalizations provided by the previous studies (see Section 1).

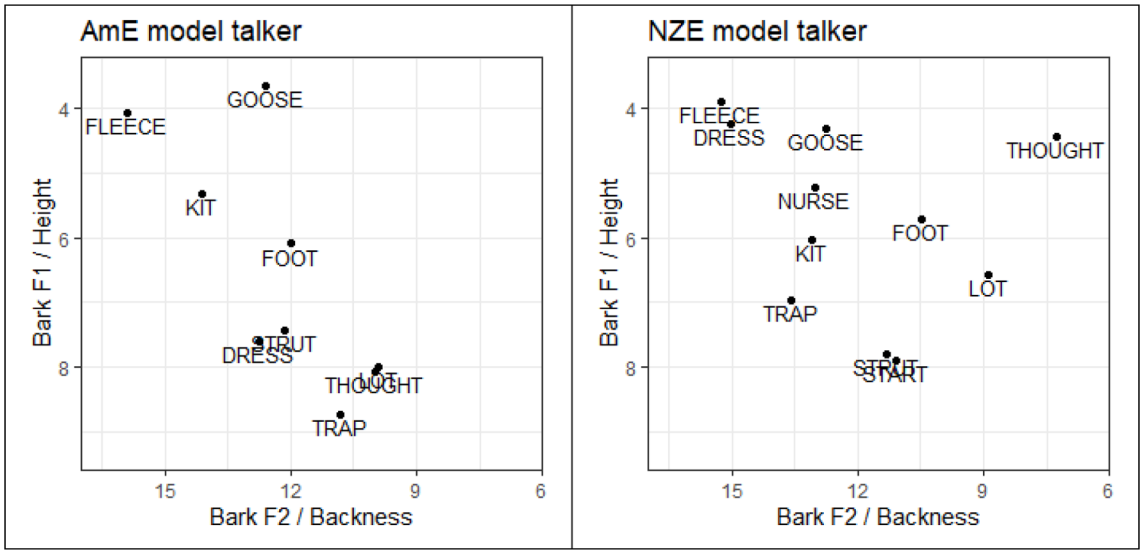

Figure 2: Model talkers' vowel spaces (Left: AmE model talker / Right: NZE model talker).

Figure 3 shows the Bark-scaled F1 values of the DRESS vowels of the model stimuli. The vertical axis indicates the Bark-scaled F1 values, and the horizontal axis expresses the model talker. Note that the *y* axis is inverted to align with vowel height. The white boxes show the distribution of the F1 values of DRESS words that do not have minimal pair competitor words with TRAP vowels, and the grey boxes show DRESS words that have minimal pair competitor words with TRAP vowels. **Table 2** provides descriptive statistics of these F1 values. As is clear, the F1 values of the model stimuli produced by the AmE model talker are higher than those produced by the NZE model talker. We fitted a mixed-effects linear regression model with F1 as the dependent variable and *dialect* (AmE vs. NZE) as a predictor. The model suggests that this difference is statistically significant ($p < .001$). This suggests that the NZE model talker

produces DRESS vowels with a higher tongue position than the AmE model talker. Once again, these descriptions of the F1 values corroborate the previous findings (see Section 1). It seems that the F1 values are slightly higher in relation to the DRESS words that have minimal pair competitor words with TRAP vowels than the DRESS words that do not have minimal pair competitor words with TRAP vowels. For each dialect, we fitted a linear regression model with F1 as the dependent variable and *minPair* (no vs. yes) as a predictor. The model suggests that the difference is statistically nonsignificant ($p > .05$), and thus there is no evidence that the F1 values of DRESS vowels in the model stimuli systematically differ depending on the existence of a minimal pair competitor word with TRAP.

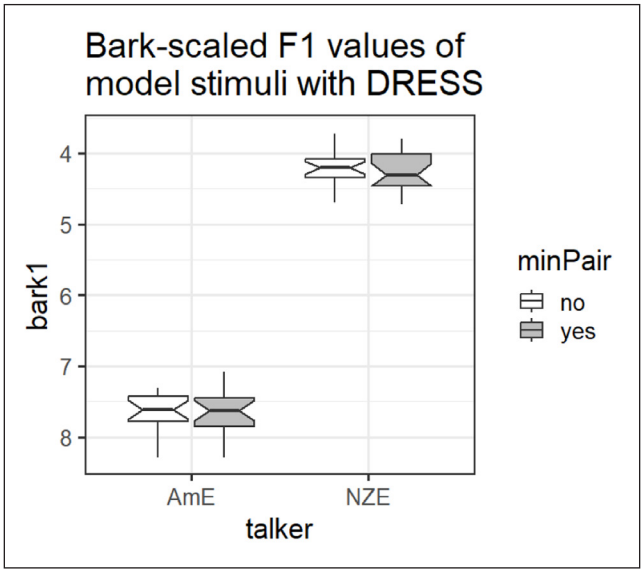

Figure 3: Distribution of F1 values of model speech stimuli.

Variable	talker: AmE	talker: NZE
minPair no	mean: 7.61, sd: 0.32	mean: 4.20, sd: 0.27
minPair yes	mean: 7.60, sd: 0.37	mean: 4.25, sd: 0.27
Overall	mean: 7.61, sd: 0.34	mean: 4.23, sd: 0.27

Table 2: Descriptive statistics of Bark-scaled F1 values of model speech stimuli.

2.3. Procedure

The shadower participants were individually invited to a sound booth in the University of Canterbury in New Zealand. The participants were assigned to either Group A or Group B. Group A participants shadowed model speech stimuli produced by the NZE model talker in the

first shadowing session and those produced by the AmE model talker in the second shadowing session. Group B participants shadowed the speech stimuli produced by the two model talkers in the reverse order. That is, the shadower participants all shadowed the two model talkers, but the order was counterbalanced between the two groups. The NZE shadowing session (control condition) was necessary because the difference between baseline and shadowed tokens (and postshadowing tokens) could be attributed to extraneous factors such as fatigue and acclimation to the experimental settings (see Martin, 2007: 196–198). To assess the true effect of the AmE shadowing (target condition), we will discuss not only the difference between the baseline and target conditions but also the difference between the baseline and control conditions in what follows. It was also necessary to counterbalance the order of the two conditions (see Martin, 2007: 156–164). Once participants are exposed to one condition, the exposure effects may last afterwards, with the result that the participants' speech behaviour may be contaminated under the other condition. In addition, participants may feel more tired in the second shadowing session than in the first shadowing session, and the difference between the two sessions could be attributed to extraneous factors. In order to minimize these order effects on the experimental results, Group A and B participants shadowed the two model talkers in the reverse order. One might wonder why we did not employ a pure between-subject design, in which each participant shadowed one model talker and no participant shadowed two model talkers. The biggest disadvantage of the pure between-subject design is that the two groups may not be equivalent to each other accidentally in terms of individual characteristics such as perceptual acuity and experience (see Martin, 2007: 151; 197–198); i.e., the difference between the two shadowing conditions might be confounded with accidental differences between the two groups. In addition, the pure between-subject design limits the number of observations that can be collected from each participant. Our experimental design allows us to collect more data, thereby further strengthening its statistical power. Even though we took extra care of the effects of session orders and individual differences, it is possible that confounding effects may still be present. This is why the statistical analyses reported in Section 3 will take the variable called *group* into consideration (see Section 2.5).

The experimental program was created using *jsPsych* (de Leeuw, 2015). The visual stimuli were presented via computer monitor (Microsoft Surface Laptop), and the audio stimuli were given over headphones (audio-technica ATH-AVC 200). The recordings were performed using a recorder (TASCAM DR-40X) and a head-mounted microphone (AKG C-520). The sampling rate was set to 44,100Hz, and the resolution was 16 bits.

In the beginning of the experiment, the participants completed a keyword reading session. They were asked to read aloud eleven keywords while seeing written forms on the PC screen: *heat*, *hid*, *head*, *had*, *hard*, *hut*, *hot*, *horse*, *hood*, *whose*, and *heard*. These words include the eleven

monophthong phonemes in NZE, and the vowel phonemes follow /h/ and precede alveolar consonants. This consonantal context is chosen to control allophonic variation (Di Paolo, 2010: 88). This initial session is supposed to familiarize the participants with the experimental settings, and the recordings in this initial session were used to capture the individual differences in the vowel spaces.

Next, they did a baseline session, and read aloud the 80 target words in **Table 1** plus the ten filler words while seeing written forms on the PC screen. The recordings of the target words were treated as baseline tokens. Next, they did the first shadowing session. The model speech stimuli were NZE tokens for Group A participants, whereas they were AmE tokens for Group B participants. They shadowed the 40 target words with DRESS vowels plus the 16 filler words with LOT or FLEECE vowels. Recall that the participants never heard the TRAP vowels from either model talker. They were not explicitly instructed to imitate the model speech stimuli, but were asked to repeat them as naturally and clearly as possible (see Babel, 2009). The prerecorded model speech stimuli were presented over headphones, and written forms appeared on the PC screen. The presentation of written forms was supposed to make sure that participants recognized which word they heard (Walker & Campbell-Kibler, 2015; Gnevsheva et al., 2021). These recordings were treated as shadowed tokens. Following this session, they did the first postshadowing session. They read aloud the 80 target words plus the ten filler words as in the baseline session. The recordings in this session were treated as postshadowing tokens. After the first shadowing and postshadowing sessions, the participants were asked to have a short break for three minutes. The purpose of this short break was to mitigate the effects of the first shadowing session.

Following the short break, they did the second shadowing session. They shadowed a different model talker from the one they shadowed in the first shadowing session. The model speech stimuli were AmE tokens for Group A participants, whereas they were NZE tokens for Group B participants. Then, they did the second postshadowing session, which is basically identical to the baseline and first postshadowing sessions. The recordings in the second shadowing and postshadowing sessions were treated as shadowed tokens and postshadowing tokens respectively. Note that the order of stimulus presentation was randomized in each of the five key sessions.

Finally, they filled in questionnaires. They answered several questions about their backgrounds (e.g., age, gender, and region where they were born and grew up). In addition, they were asked which dialect the first and second model talkers spoke, and why they thought so. According to the questionnaire data, 17 of the 24 participants correctly identified the AmE model talker as an AmE speaker, five of them identified her as a Canadian English speaker, and the other two participants identified her as a British English speaker. This suggests that the majority of the

shadower participants may raise the activation of "North American English" at least. As for the NZE model talker, all the participants correctly identified her as a NZE speaker. The experimental procedure can be summarized as in **Table 3**.

	Group A (NZE_first)	**Group B (AmE_first)**
1.	Keyword reading session	
2.	Baseline session	
3.	NZE shadowing session	AmE shadowing session
4.	Postshadowing session I	
THREE-MINUTE SHORT BREAK		
5.	AmE shadowing session	NZE shadowing session
6.	Postshadowing session II	
7.	Questionnaires	

Table 3: Experimental procedure.

2.4. Acoustic analysis and response variable

Our interest lies in the formant frequency values of DRESS and TRAP vowels. We manually annotated the vocalic domains of the target word tokens in the five key sessions using Praat (Boersma & Weenink, 2020). We defined the vocalic domain by taking into consideration the periodicity and intensity of waveform. In addition, we annotated the eleven monophthongs recorded in the keyword reading session. Using a Praat script, the formant measurements were performed. In particular, the central 50% of the vocalic parts were extracted using a Hamming window, and the average formant values of the extracted regions were measured with five formants tracked and the maximum frequency set to 5,500 Hz.

Following this, the raw frequencies were converted into Bark (Traunmüller, 1990). It is well known that human auditory systems are more sensitive to differences in low frequency sounds than high frequency sounds (Johnson, 2012: 88). The Bark scale reflects the fact that the sensation of lower frequency sounds is more minutely divided than that of higher frequency sounds. Bark scaled F1 values were fitted as a response variable in the statistical models reported in Section 3. This variable is called *Bark1* in what follows.[3]

[3] Note that our initial plan was to explore F2 values as well as F1 values when we preregistered the current study. However, we decided to report only the results with respect to F1 values in this research article. As we explored our dataset, we found that DRESS vowels vary between NZE and AmE specifically in terms of F1 values, and so do DRESS and TRAP vowels in NZE (see Section 3.1). In fact, Sanchez et al. (2015) also note that DRESS and TRAP in NZE are distinct particularly in F1 values. Although the current study does not discuss the variability in F2 values, the shared data include F2 values, which were measured in the same way as illustrated in Section 3.4, in the spirit of open science.

2.5. Explanatory variables

Our general interest lies in whether NZE speakers adjust their production of DRESS to the AmE model talker during the shadowing session, and whether the adjustment is observed after the shadowing session, that is, during the postshadowing session. Furthermore, a key question is whether the convergence effects can be extended to the unheard vowel, TRAP, in the postshadowing session. In order to explore these convergence effects, we coded our dataset depending on the experimental sessions (see **Table 3**). The tokens produced in the baseline session were coded as baseline tokens. Those in the AmE shadowing session were coded as AmE shadowing tokens, and those in the following postshadowing session were coded as AmE postshadowing tokens. Likewise, the tokens produced in the NZE shadowing session were coded as NZE shadowing tokens, and those produced in the following postshadowing session were coded as NZE postshadowing tokens. This multi-level variable is called *task* in the following statistical analyses, and fitted as a key variable. Note that this variable is a five-level variable for DRESS vowels ("baseline", "AmE shadowing", "AmE postshadowing", "NZE shadowing", and "NZE postshadowing"), while it is a three-level variable for TRAP vowels ("baseline", "AmE postshadowing", and "NZE postshadowing"). This is because TRAP words were not employed in the shadowing sessions. This variable is treatment-coded, and the reference level is "baseline". If the participants converged towards the model talkers, statistical significance would be expected for this variable.

Our key interest also lies in whether the imitation of a word with DRESS is influenced by the presence of a minimal pair competitor word with TRAP. To address this research question, we coded a token in our dataset as "yes" if the word has a minimal pair competitor with the other target vowel. For example, the tokens of *bad* and *bed* were coded as "yes", whereas the tokens of *fact* and *deaf* were coded as "no". This binary variable is called *minPair* in what follows, and fitted as a key variable. This variable is treatment-coded, and the reference level is "no". If the extent of phonetic convergence were influenced by the presence of a minimal pair competitor, the interaction between *minPair* and *task* should reach statistical significance.

The tokens were also coded depending on the group. The tokens produced by Group A participants were coded as "NZE_first", whereas those produced by Group B participants were coded as "AmE_first". The reason why this variable is fitted as an explanatory variable in a statistical model is because the first exposure to a particular accent may influence speech production throughout the experiment (see Section 2.3). This binary variable is called *group* in the following statistical analyses, and fitted as a control variable. This variable is treatment-coded, and the reference level is "NZE_first". If the exposure effect in the first shadowing session

persisted until the end of the experiment, statistical significance would be expected for the interaction between *group* and *task*.[4]

2.6. Data removal

Recall that the 24 NZE speakers read aloud the 80 target words with DRESS and TRAP vowels in the baseline session and the two postshadowing sessions, and shadowed the model speech stimuli of the 40 target words with DRESS vowels in the two shadowing sessions (see **Table 1** for the list of target words). With respect to target words, each speaker produced 80 baseline tokens, 80 shadowed tokens, and 160 postshadowing tokens. Hence, we aimed to collect 1,920 baseline tokens (960 DRESS vowel tokens and 960 TRAP vowel tokens), 1,920 shadowed tokens (1,920 DRESS vowel tokens), and 3,840 postshadowing tokens (1,920 DRESS vowel tokens and 1,920 TRAP vowel tokens) in total.

Some tokens were removed from the following statistical analyses. First, we examined the extent to which each participant's F1 values for the DRESS and TRAP vowels changed during the shadowing and postshadowing sessions compared to the baseline session. We found that one participant was more than 3 SD from the ground mean value in several sessions, indicating that they behaved in a very different manner from the rest of the participants. This participant was therefore removed from the statistical analysis.

Second, a token was removed if its F1 frequency value deviated by 3 standard deviations from the mean F1 value of an individual participant for each of the two target vowels within a session. In the end, 7,299 tokens remained in our dataset. The statistical analyses were performed on the 7,299 tokens in what follows.[5]

3. Statistical analysis

3.1. Descriptive statistics

Before proceeding to the main statistical analyses, we present descriptive statistics to explore patterns in the vowel productions of the shadower participants. **Figure 4** shows the grand mean formant values of the 11 NZE monophthongs (KIT, DRESS, TRAP, STRUT, LOT, FOOT, FLEECE, NURSE, START, THOUGHT, and GOOSE) across the shadower participants in the keyword reading session. The scales of the horizontal and vertical axes of **Figure 4** are identical to those of **Figure**

[4] When we preregistered this study, we planned to explore order (i.e., trial number) within a session as a control variable. We anticipated that speech production may be influenced by fatigue and acclimation over the course of a session, and the originally submitted analysis included a significant order that did not substantially affect the main conclusion. Following the advice of editor and anonymous reviewers, we simplified the analysis. The information of order is available in supplementary materials for anyone interested.

[5] The second removal criterion was preregistered. The first one was added following data exploration.

2. It exhibits some well-known characteristics of the NZE vowel system, such as DRESS raising, KIT centralization, and GOOSE fronting, suggesting that the shadower participants generally had a typical NZE vowel system. A comparison of the shadower participants' mean vowel space and the NZE model talker's vowel space (shown in the right panel of **Figure 2**) reveals a similar vowel distribution, suggesting that the NZE model talker is also a typical NZE speaker. The figure also indicates that the target vowels, DRESS and TRAP, are distinguished primarily by F1 values.

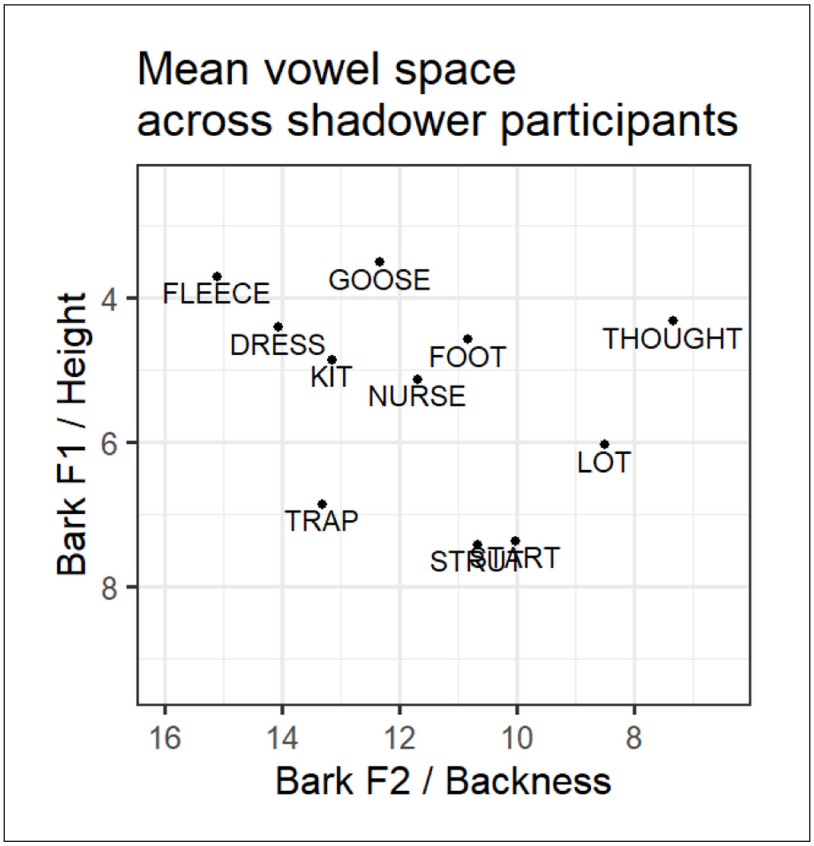

Figure 4: Mean vowel spaces of shadower participants based on keyword tokens.

Next, we would like to examine whether the DRESS vowels of the shadower participants tend to be higher in the vowel space compared to those of the AmE model talker, and whether the TRAP vowels of the shadower participants tend to be close to the DRESS vowels of the AmE model talker. **Figure 5** compares the formant values of the shadower participants and the model talkers. The green dot expresses the mean value of the DRESS vowels in the NZE model speech stimuli, and the red dot indicates the mean value of the DRESS vowels in the AmE model speech stimuli. The individual black dots in the left-hand figure express the mean values of the DRESS vowels of the shadower participants in the baseline session, and those in the right-hand figure

represent the mean values of the TRAP vowels of the shadower participants in the baseline session. The left-hand figure clearly shows that the DRESS vowels of shadower participants are higher than those of the AmE model talker in vowel space. The right-hand figure suggests that the majority of the shadower participants produced TRAP vowels with formant values that are close to the DRESS vowel of the AmE model talker. These findings corroborate the previous studies (Gordon et al., 2004; Hay et al., 2008, 2015; Ladefoged & Johnson, 2015), and suggest that our data is well suited for addressing the research questions outlined in Section 1.

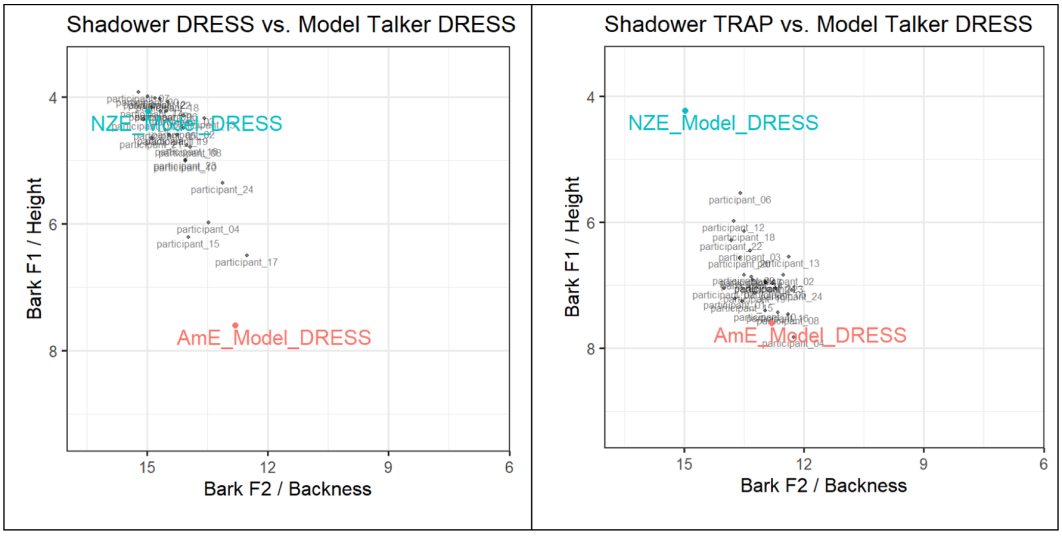

Figure 5: Baseline formant values of shadower participants and model talkers.

Finally, we would like to explore the F1 values of DRESS and TRAP vowels in each key session. The top panel of **Figure 6** visualizes the distribution of F1 values for DRESS, and the bottom panel shows the distribution for TRAP. The *x* axis expresses *task*, and the *y* axis shows Bark-scaled F1 values. The red horizontal line represents the mean Bark F1 value of the AmE target model stimuli (mean: 7.61), while the green line represents that of the NZE target model stimuli (mean: 4.23). Recall that the shadower participants never shadowed the TRAP vowels from either model talker, and the target model stimuli included only DRESS vowels. **Table 4** shows the descriptive statistical values. Clearly, the production of DRESS shifted towards AmE model speech stimuli in the AmE shadowing task. It is worth noting that the mean DRESS F1 value in the AmE shadowing session is still far from that of the AmE model stimuli, suggesting that shadower participants did not completely align their production with an AmE model talker. The other potential changes made in the shadowing and postshadowing tasks are not so clear from these descriptive statistics, and they are further explored using mixed-effects linear regression analyses in the next section.

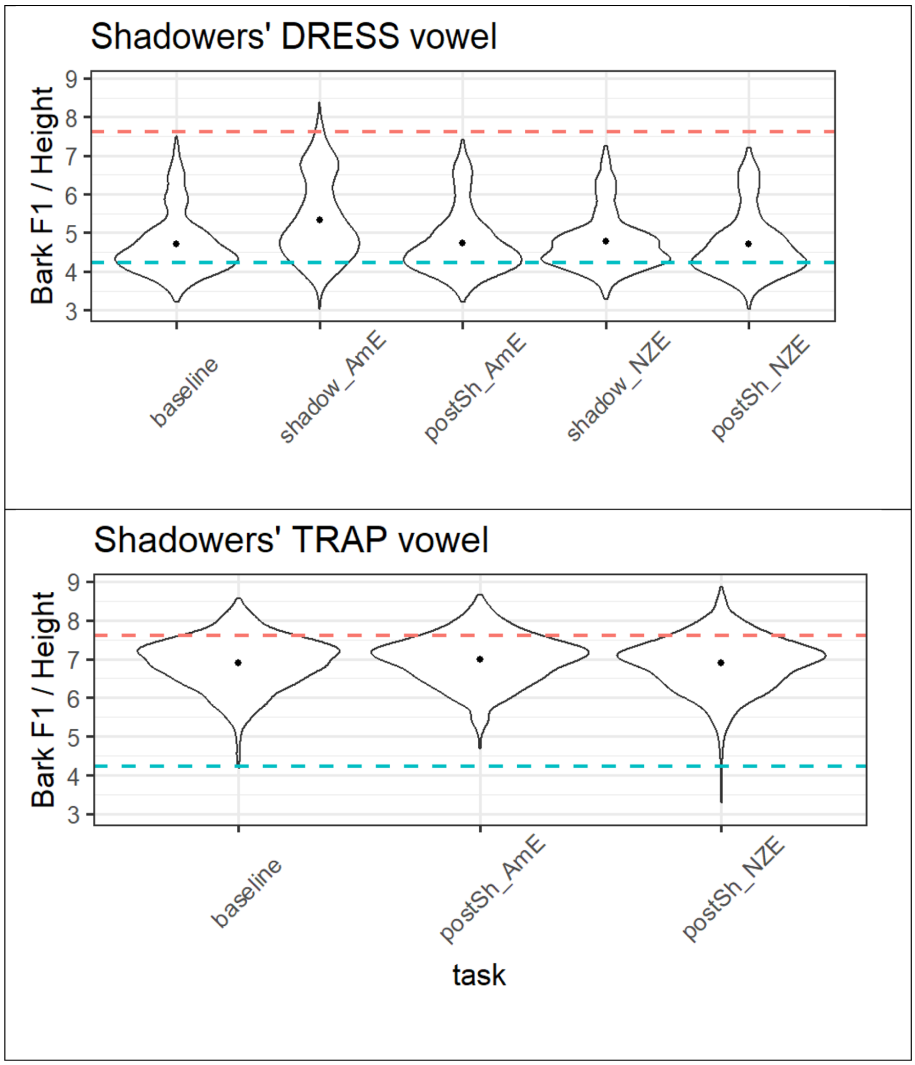

Figure 6: Violin plot of Bark F1 value in each key task. The dots show mean Bark F1 values. The red horizontal line shows the mean Bark F1 value of AmE model DRESS stimuli, and the green line shows that of NZE model DRESS stimuli.

task	DRESS	TRAP
Baseline	mean: 4.7, sd: 0.79	mean: 6.89, sd: 0.66
AmE shadowing	mean: 5.33, sd: 1.03	
AmE postshadowing	mean: 4.73, sd: 0.82	mean: 6.98, sd: 0.63
NZE shadowing	mean: 4.76, sd: 0.73	
NZE postshadowing	mean: 4.7, sd: 0.82	mean: 6.88, sd: 0.7

Table 4: Descriptive statistics of shadowers' Bark-scaled F1 values

3.2. Mixed-effects linear regression analysis

The statistical analyses were performed using the *lme4* library 1.1-35 (Bates et al., 2015) implemented in R (R Core Team, 2024) with the aid of the optimizer *optimx* 10.21 (Nash, 2014). As discussed in the preceding section, the response variable is a bark-scaled F1 value (*Bark1*), and the explanatory variables are *task*, *minPair*, and *group*. The DRESS and TRAP vowels were analysed separately. First, we constructed the maximal model with the three-way interaction of the predictor variables, by-participant and by-word random intercepts. Then, backward elimination was performed iteratively, manually removing variables one by one. In the course of the backward elimination, we incorporated by-participant random slopes for the key variables so that the models would converge reliably. The elimination process was based on *p* values, which were obtained by comparing a fuller model with a reduced model lacking a particular variable or interaction term. Model comparisons were done using *anova*. If a model comparison shows no significance ($p > .05$), then the reduced model was adopted; otherwise, the fuller model was adopted. Following this procedure, the two models in (3) were selected as the best-fitted models. The best-fitted models did not have any statistical convergence errors. With respect to the DRESS model, the variable *group* was removed from further analysis because its main effect and interaction terms with other predictors were not significant ($p > .05$). With regards to the TRAP model, the variables *group* and *minPair* were removed from further analysis because their main effects and interaction terms with other predictors did not improve the model ($p > .05$). VIF values were checked to avoid overfitting problems. All VIF values are lower than 4, suggesting that there is no multicollinearity in the best fitted models. The model summaries are provided in **Tables 5** and **6**. As the *lme4* library does not provide *p* values, we used *lmerTest* library 3.1-3 (Kuznetsova et al., 2017) to retrieve the *p* values in the tables.

(3) *lmer* syntax of best fitted model

a. DRESS model

$$Bark1_{DRESS} \sim 1 + task * minPair + (1 + task + minPair \mid participant) + (1 \mid word)$$

b. TRAP model

$$Bark1_{TRAP} \sim 1 + task + (1 + task \mid participant) + (1 \mid word)$$

Together, these two best fitted models can answer our research questions. We will outline these results and our background assumption (that DRESS would be affected by AmE), and our two research questions in separate subsections. Recall that the variable *group* was removed through the backward elimination due to its lack of statistical significance. The statistical nonsignificance suggests that the effects discussed below were observed regardless of the assigned group.

3.2.1. The realization of DRESS in shadowed and nonshadowed conditions

First of all, let us discuss the direct effect of the shadowing. We assumed that the speakers' DRESS vowels would shift towards the model's DRESS vowels, and that this effect may still be visible

in the postshadowing task. The DRESS model (**Table 5**) allows us to explore these shadowing effects. The interaction between *task* and *minPair* significantly improved the model ($p = .0036$). Note that the p value was retrieved by comparing the best fitted model (3a) and the reduced model lacking the interaction term. This significant interaction suggests that the difference between the baseline and other sessions depends on the presence of a minimal pair competitor containing a TRAP vowel. In order to assess the true effects of *task* on $Bark1_{DRESS}$, we performed a post hoc analysis. The variable *minPair* was sum-coded, and the DRESS model was rerun. **Table 7** is the summary table of the model with sum-coded *minPair*. The sum-coded model allows us to explore the effect of *task* while *minPair* is centred; i.e., the intercept shows the estimated F1 value when *task* is baseline and *minPair* is the unweighted grand mean value (Sonderegger 2023: 202). First, let us discuss the target conditions (i.e., the AmE shadowing and postshadowing sessions). Unsurprisingly, the difference between the baseline and AmE shadowing sessions is significant (Estimate (0.62), Std. Error (0.20), t-value (3.00), and $p = .006$). The effects are in the expected direction; i.e., the shadower participants produced DRESS vowels with more AmE-like formant values (i.e., higher F1 values) in the shadowing session than in the baseline session. The difference between the baseline and AmE postshadowing sessions is non-significant (Estimate: 0.03, Std. Error: 0.02, t value: 3.00, and $p = .14$), although the effects are in the expected direction. Next, let us discuss the control conditions (i.e., the NZE shadowing and postshadowing sessions). The difference between the baseline and the NZE shadowing session was not significant, nor was the difference between the baseline and the NZE postshadowing sessions ($ps > .05$). That is, there is no robust evidence that the production of DRESS in the control conditions differed from that in the baseline task. In sum, participants shifted their DRESS vowels towards the AmE model during the shadowing task, but this shift did not persist into the AmE postshadowing session.

Variable	Estimate	Std. Error	*t* value	*p* value
(Intercept)	4.705	0.1549	30.36	$<.001$
task shadow_AmE	0.5814	0.21	2.76	0.0109
task postSh_AmE	0.0186	0.0269	0.69	0.49
task shadow_NZE	0.0029	0.0393	0.742	0.46
task postSh_NZE	0.0007	0.0257	0.029	0.97
minPair yes	0.0031	0.062	0.05	0.96
task shadow_AmE: *minPair* yes	0.0947	0.0298	3.176	0.001
task postSh_AmE: *minPair* yes	0.0313	0.0297	1.055	0.29
task shadow_NZE: *minPair* yes	0.0737	0.0297	2.476	0.01
task postSh_NZE: *minPair* yes	0.0065	0.0297	0.219	0.82

Table 5: Summary of DRESS model. The reference level is set as *task* baseline and *minPair* no.

Variable	Estimate	Std. Error	*t* value	*p* value
(Intercept)	6.8973	0.1111	62.053	<.001
task postSh_AmE	0.0911	0.0329	2.766	0.01
task postSh_NZE	−0.0125	0.0358	−0.349	0.703

Table 6: Summary of TRAP model. The reference level is set as *task* baseline.

Variable	Estimate	Std. Error	*t* value	Pr(>|t|)
(Intercept)	4.706	0.1513	31.105	< 2e–16
task shadow_AmE	0.6287	0.2095	3.001	0.006
task postSh_AmE	0.0343	0.0225	1.524	0.14
task shadow_NZE	0.066	0.0364	1.812	0.083
task postSh_NZE	0.004	0.021	0.19	0.85
minPair[S.no]	−0.0015	0.031	−0.05	0.96
task shadow_AmE: *minPair*[S.no]	−0.0473	0.0149	−3.176	0.001
task postSh_AmE: *minPair*[S.no]	−0.0156	0.0148	−1.055	0.29
task shadow_NZE: *minPair*[S.no]	−0.0368	0.0148	−2.476	0.01
task postSh_NZE: *minPair*[S.no]	−0.0032	0.0148	−0.219	0.82

Table 7: Summary tables of $Bark1_{DRESS} \sim task * minPair$ (sum-coded) + (1 + *task* + *minPair* | *participant*) + (1 | *word*). The reference level is set as *task* baseline.

The lack of a significant difference between the baseline and postshadowing sessions may be due to participants returning to their baseline production patterns after a certain number of trials following the shadowing session. To test this possibility, we analysed only the first halves of the postshadowing sessions. These subset analyses revealed a significant difference between the baseline and AmE postshadowing sessions with respect to the first halves (Estimate (0.064), Std. Error (0.024), t-value (2.64), and $p = .01$), whereas the difference between the baseline and NZE postshadowing sessions was non-significant (Estimate (0.025), Std. Error (0.022), t-value (1.11), and $p = .26$). Note that these differences are non-significant in the second halves of the postshadowing sessions ($ps > .05$). That is, the shift towards the model's DRESS persisted only in the early AmE postshadowing session. The difference between the baseline and AmE postshadowing sessions gradually decreased as the session progressed.

3.2.2. The realization of TRAP after shadowing DRESS (RQ1)

In research question 1, we asked whether the social activation invoked by exposure to American English would have an effect on other vowels. Specifically, we asked whether

TRAP would be affected, despite there being no within-task exposure to AmE TRAP. This would be shown by an effect of task in the TRAP model (**Table 6**). The difference between the baseline and AmE postshadowing sessions is significantly different in the predicted direction: Estimate: 0.09, Std. Error: 0.11, *t*-value: 2.76, and $p = .01$. As for the control condition (i.e., the NZE postshadowing session), its difference from the baseline session is not significant at all ($p = 0.73$). To sum up, the shadower participants produced TRAP vowels with more AmE-like formant values (i.e., higher F1 values) in the AmE postshadowing session than in the baseline session; however, no such carryover effects were observed in the NZE postshadowing session.

3.2.3. The effect of TRAP minimal pairs on DRESS (RQ2)

Finally, in research question 2, we asked whether there would be greater shadowing effect for DRESS words which did not have TRAP minimal pair words. This would be demonstrated by a significant effect of *minPair* in the DRESS model. We see that the DRESS model (**Table 5**) contains a significant interaction between *task* and *minPair* ($p = 0.0036$). This significant interaction suggests that the difference between the DRESS words with minimal pair competitor words containing TRAP vowels and those without such competitor words varies depending on *task*. **Figure 7** visualizes the interaction between *minPair* and *task* in relation to $Bark1_{DRESS}$. The *y* axis represents mean $Bark1_{DRESS}$, the *x* axis represents *task*, and the fill colour represents *minPair*. Dots represent mean values, and error bars represent standard error. The red horizontal line shows the mean Bark F1 value of AmE model DRESS stimuli (mean: 7.61), and the green line shows that of NZE model DRESS stimuli (mean: 4.23) as in **Figure 6**. In this figure, the difference between the two types of words in the F1 values of the DRESS vowel looks larger in the AmE and NZE shadowing sessions than in the baseline and postshadowing sessions, and the F1 value of a DRESS vowel appears to be higher if the word has a minimal pair competitor with a TRAP vowel, particularly in the AmE and NZE shadowing sessions. In order to assess whether these differences are statistically significant within each session, we split the dataset into the five tasks, and fitted a mixed-effects linear regression of *minPair* on $Bark1_{DRESS}$ to each subset. The summary tables are shown in **Table 8**. It turned out that the difference between words with minimal pair competitor words containing TRAP vowels and those without such competitor words is not significant within any of the five sessions ($p > .05$). The results of these post hoc analyses suggest that the difference between the two types of words is modulated by *task*, but they are not modulated enough to change the significance of the *minPair* effect for any subset of tasks (see Sonderegger, 2023, 223). That is, there was no robust evidence that the imitation of AmE DRESS vowels is influenced by the presence of a minimal pair competitor with TRAP vowels.

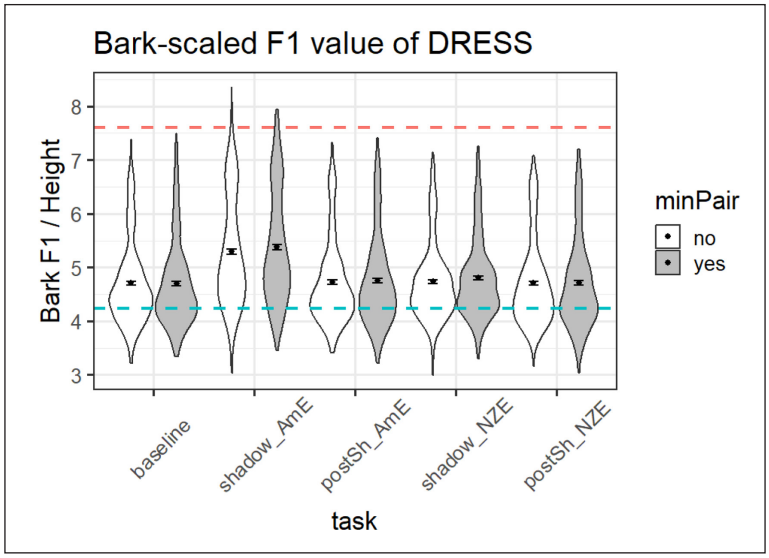

Figure 7: Violin plot of $Bark1_{DRESS}$ value as a function of *task* and *minPair*. The dots show mean Bark F1 values, and error bars represent standard error. The red horizontal line shows the mean Bark F1 value of AmE model DRESS stimuli, and the green line shows that of NZE model DRESS stimuli.

Variable	Estimate	Std. Error	*t* value	*p* value
Baseline				
(Intercept)	4.7049	0.1537	30.594	< 2e–16
minPair yes	0.0035	0.0576	0.062	0.951
AmE shadowing				
(Intercept)	5.2864	0.2025	26.104	< 2e–16
minPair yes	0.0976	0.0594	1.642	0.109
AmE postshadowing				
(Intercept)	4.7236	0.1606	29.396	< 2e–16
minPair yes	0.0342	0.0644	0.532	0.598
NZE shadowing				
(Intercept)	4.7334	0.1442	32.81	< 2e–16
minPair yes	0.0775	0.0614	1.263	0.214
NZE postshadowing				
(Intercept)	4.7054	0.1608	29.251	< 2e–16
minPair yes	0.009	0.0587	0.154	0.879

Table 8: Summary tables of $Bark1_{DRESS} \sim minPair + (1 + minPair \mid participant) + (1 \mid word)$ for each task. The reference level is set as *minPair* no.

4. Discussion and summary

The current study employed cross-dialect convergence as a lens to learn things about the mental representations and cognitive mechanisms that are involved in phonetic imitation. In order to investigate this broad question, the current study addressed the two key questions in relation to cross-dialect convergence. Section 4.1 summarizes the general findings on cross-dialect convergence, and Section 4.2 discusses the key findings in relation to the two key research questions.

4.1. General findings on cross-dialect convergence

The current study contributed another piece of empirical evidence for cross-dialect convergence. As reviewed in Section 1.1, the previous studies demonstrated that a shadower participant shifts their phonetic articulation in the direction of the model speech, even though the shadower participant and model talker have different dialectal backgrounds (Delvaux & Soquet, 2007; Babel, 2010; Dufour & Nguyen, 2013; German et al., 2013; Walker & Campbell-Kibler, 2015; Gnevsheva et al., 2021; Clopper et al., 2024; Lee-Kim & Chou, 2024). Our experimental results demonstrated that NZE speakers modified the production of DRESS vowels while and after shadowing the model speech stimuli including AmE DRESS vowels. The NZE shadower participants produced DRESS vowels with lower F1 values in the baseline sessions as compared with the AmE model talker (see **Figure 5**). It was demonstrated that they produced DRESS vowels with significantly higher F1 values in the AmE shadowing session than in the baseline session (see **Table 5**). In addition, the shadower participants produced DRESS vowels with significantly higher F1 values in the early AmE postshadowing session than in the baseline session (see Section 3.2.1). The shadower participants did not exhibit such shifts in the NZE shadowing and postshadowing sessions, suggesting that the shifts towards AmE DRESS vowels in the AmE shadowing and postshadowing sessions cannot be attributed to extraneous factors such as fatigue and experimental acclimation.

Exemplar Theory provides a straightforward account of these cross-dialect convergence effects (see Section 1.1). When a NZE speaker hears AmE model speech stimuli, the auditory experiences may be stored as new exemplars in memory. The newly stored exemplars have relatively high resting activation levels due to the recency of memory storage (Pierrehumbert 2001; Wedel 2006). Hence, the recently experienced AmE DRESS vowels are likely to influence the creation of the production target in the AmE shadowing and postshadowing tasks. Consequently, the shadowed and postshadowing utterances are likely to have phonetic features that are similar to the speech stimuli that are produced by the AmE model talker (e.g., higher F1 values with respect to DRESS vowels).

The observed cross-dialect convergence may be attributed not only to this bottom-up processing, but also to top-down mechanisms. In particular, social category activation may also contribute to the convergence effects observed for DRESS. Exposure to AmE model speech

stimuli may raise the activation of the relevant social category "American English," which in turn increase the activation levels of the preexisting exemplars that encode AmE DRESS vowels. Consequently, exemplars encoding AmE DRESS may become more readily available when speakers form a production target (see Section 4.2.1).

The best fitted model (**Table 5**) suggests that the convergence effects are larger in the AmE shadowing session than in the AmE postshadowing session. The estimated difference in Bark-scaled F1 between the baseline and AmE shadowing sessions is about 0.581, whereas the difference between the baseline and AmE postshadowing sessions is about 0.018. That is, the shifts towards AmE DRESS vowels were attenuated in the AmE postshadowing session. This attenuation of the convergence effects can be interpreted to mean that the resting activation levels of newly stored exemplars gradually decreased over time since they were stored. Newly stored exemplars may have a strong influence on the formation of a production target at the time of storage, but this influence may diminish over time. Once the activation levels of these exemplars drop significantly, speech production is likely to rely more on the preexisting exemplars that habitually contribute to the creation of a production target in daily speech. Pierrehumbert (2001) refers to this process as "memory decay," wherein the resting activation levels of exemplars gradually decrease over time. Wedel (2006) demonstrates that this process plays an important role in sound change. Our findings provide further empirical evidence consistent with memory decay.

According to our statistical analyses, there is no robust evidence that the convergence effects differed between Groups A and B. As discussed in Section 3.2, the variable *group* did not improve the statistical models, as neither its main effects nor its interactions with other predictors were significant. This can be interpreted to mean that the two groups were comparable, and the order of dialect exposure did not affect the extent of phonetic imitation. This finding is consistent with Walker & Campbell-Kibler (2015). In their experiment, the participants shadowed model speech produced by speakers of NZE, Australian English, US Midland English, and US Inland North English. Some participants shadowed the model talkers in this order, while others shadowed them in the reverse order. With respect to DRESS vowels, the participants in the experiment converged towards the AmE model talkers regardless of the order in which the model talkers were presented. The absence of exposure-order effects may suggest that newly stored exemplars are fragile and susceptible to rapid memory decay. Once a shadower is exposed to a new model talker, the exemplars encoding the speech of the previous model talker may quickly become less accessible.[6]

The best fitted model (**Table 5**) suggests that the shadower participants exhibited a relatively small degree of convergence towards the AmE model talker. The estimated

[6] Relatedly, Kraljic & Samuel (2007) show that the most recently heard utterances override previously heard utterances in the perceptual adjustment of /t/ versus /d/.

Bark-scaled F1 value of DRESS in the baseline session is approximately 4.69, whereas in the AmE shadowing session, it is about 5.26; i.e., the shift is only 0.57. The mean F1 value of the AmE model speech stimuli was 7.61 (see **Table 2**). That is, the difference between the AmE shadowed tokens and model speech stimuli was approximately 2.35, suggesting that shadower participants had room for further phonetic imitation. A similar finding was also reported by Walker & Campbell-Kibler (2015). In their experiment, NZE speakers shadowed AmE model speech stimuli with various vowel phonemes. The NZE speakers shifted their DRESS vowels towards the AmE model talker by around 0.6 in Bark-scaled F1, although there remained room for further phonetic imitation. The limited extent of imitation is not unique to cross-dialect convergence but has also been observed in within-dialect convergence. Babel (2010) investigated the extent to which speakers of Californian English converge towards Californian English model talkers. As in the current study, the participants shadowed model speech stimuli. The vowel plots in their study showed that shadowed utterances moved closer to the model talker's formant values, although a measurable distance between the two remained. Nielsen (2011) also investigated within-dialect convergence. In their Experiment 1, AmE speakers were exposed to words containing lengthened /p/. The mean VOT value of the lengthened /p/ was about 113 ms. The AmE speakers produced /p/ with about 66 ms in the preexposure session and approximately 73 ms in the postexposure session, indicating substantial room for convergence.

Why phonetic imitation is relatively limited, and why ample room for further imitation remains, may be attributable to several factors. One possible explanation comes from Exemplar Theory. Under this account, multiple exemplars are selected and averaged to form a production target (see Section 1.1). Although a recently heard utterance may exert a relatively strong influence on target formation, exemplars that habitually contribute to the production target continue to shape a production target. As a result, shadowed and postshadowing utterances may reflect both the model stimuli and speakers' habitual productions, thereby limiting the extent of convergence towards the model talker. An alternative explanation concerns constraints imposed by the articulatory motor system. Even when speakers attempt to shift their productions towards a model talker, articulatory constraints may limit the extent of such adjustments. Supporting this view, previous research has shown that L2 learners often struggle to produce nonnative sounds, even when explicitly instructed on articulatory movements (Saito, 2011). In the case of cross-dialect convergence, social constraints may also play a role. Sociolinguistic variation conveys social meanings associated with particular groups and values (Bell, 2014, 270–271; Eckert, 2016). Shadowers may therefore avoid producing variants that deviate too strongly from their habitual speech patterns, as such variants may convey undesirable social messages.

4.2. Key research questions

4.2.1. Effect of shadowing a vowel phoneme from a different dialect on the postexposure production of another vowel phoneme

The first key research question was whether imitation of a particular vowel indexed to a different dialect is systematically generalized to another vowel phoneme that is likewise indexed to that dialect. To address this research question, we explored whether NZE speakers modify the production of TRAP vowels after shadowing AmE DRESS vowels. They never heard AmE TRAP vowels in the shadowing session. It was found that the participants produced TRAP vowels with higher F1 values (i.e., more AmE-like formant values) in the AmE postshadowing session than in the baseline session. They did not exhibit such shifts in the NZE postshadowing session, suggesting that the shifts towards AmE TRAP vowels cannot be attributed to extraneous factors such as fatigue and experimental acclimation.

The observed TRAP shift is in line with our prediction, and supports the hypothesis that spreading activation occurs between sociolinguistic categories and associated exemplars. As NZE speakers are frequently exposed to AmE speech through popular media (Walker & Campbell-Kibler, 2015), they may memorize some speech experiences with AmE TRAP vowels. Listening to the model speech stimuli produced by the AmE model talker may raise the activation of the relevant social concept "American English." The activation of the social category would increase the activation levels of the preexisting exemplars that encode AmE speech. Hence, in the AmE postshadowing session, the exemplar memories that encode AmE TRAP vowels may become more readily available, and more likely to contribute to the creation of a production target. In consequence, TRAP vowels were produced with more AmE-like formant values. The current study contributes another piece of evidence to the growing body of literature documenting the dialect priming effects on speech (Hay et al., 2006; Love & Walker, 2013; Sanchez et al., 2015; Hashimoto, 2019; Wade, 2022). We provided additional evidence that exposure to a different dialect may have a broad influence on postexposure speech production. The current study contributes to a growing body of research demonstrating that sociolinguistic categories are mentally represented and linked to linguistic knowledge.

As TRAP and DRESS are neighbouring vowels, one might wonder whether the modification of TRAP is caused by intelligibility biases at the phonemic level (i.e., biases towards preserving phonemic contrasts) rather than spreading activation between sociolinguistic categories and associated exemplars. As discussed in Section 1.2, a speaker is hypothesized to ensure that a speech signal is intelligible enough to be understood by others. Our results demonstrate that the NZE shadower participants produced DRESS vowels with higher F1 values in the early AmE postshadowing session (see Section 4.1). As a result, the pronunciation of DRESS vowels became slightly similar to their habitual pronunciation of TRAP vowels. NZE speakers may be biased to change the pronunciations of TRAP vowels accordingly to maintain the phonemic distinction

between the two vowel phonemes, with the result that they might become likely to produce TRAP vowels with higher F1 values than usual.

On balance, our results seem more compatible with the social-category-activation account rather than the intelligibility-bias account. There are two reasons for this speculation. The first evidence comes from the extent of the modification of DRESS vowels in the AmE shadowing and postshadowing sessions. The shadower participants' shifts towards AmE DRESS vowels may not have caused sufficient phonemic ambiguity between the DRESS and TRAP vowels. According to the best fitted model, the estimated Bark-scaled F1 value of DRESS in the baseline session is approximately 4.705, and that in the early AmE postshadowing session is about 4.773. That is, the difference between the two sessions is roughly 0.067. This adjustment may not be sufficient to lead to the phonemic ambiguity between DRESS and TRAP vowels. **Figure 4** shows the mean vowel space across the shadower participants on the basis of the keyword reading session. The difference between the two vowels in Bark-scaled F1 is approximately 2.5 on average, which is much larger than the difference between the baseline and AmE postshadowing sessions in Bark-scaled F1 values of DRESS. This suggests that, even though the production of DRESS shifted towards AmE-like pronunciations in the AmE postshadowing session, the formant values of DRESS were still far away from those of TRAP in their habitual production. Furthermore, the best fitted model for TRAP F1 values (**Table 6**) indicates that the overall difference between the baseline and AmE postshadowing sessions in Bark-scaled F1 values of TRAP is about 0.09, which is also quite small as compared with the distance between DRESS and TRAP vowels in their habitual production. The intelligibility-based accounts may not yield clear predictions regarding these relatively slight shifts of TRAP and DRESS vowels in the AmE postshadowing session, as such small changes are unlikely to affect phonemic intelligibility.

These relatively slight shifts of TRAP vowels in the AmE postshadowing sessions are compatible with the social-category-activation account. As discussed above, the activation of the social concept "American English" may increase the activation levels of the exemplars that encode AmE TRAP vowels. This does not necessarily mean that the production target is formed solely on the basis of the exemplars that encode AmE vowels afterwards. Exemplars that encode NZE speech should still contribute to the creation of the production target dominantly. This is because they habitually contribute to the creation of a production target in daily speech, and have potentially high activation levels. The slight shifts of the target vowels naturally occur given that the production target is formed based on the dominant number of exemplars that encode NZE vowels and the small number of exemplars that encode AmE vowels. In fact, the previous studies that report the effects of social-category-activation also show that the shifts in formant values are very small. Love & Walker (2013) show that English speakers living in the US lower the F3 values of /ɹ/ when they talk about American football as compared with when they talk about English soccer. According to the statistical model reported in their study, the shift is only

0.058 in Bark scale. Sanchez et al. (2015) demonstrate that NZE speakers decreased F1 values in the production of KIT vowels when their speech was related to Australia in comparison with when their speech was related to something neutral. The change is only 0.096 in Bark-scaled F1 values in accordance with the statistical model reported in their study.

The second evidence comes from a post hoc comparison of TRAP and DRESS with respect to the changes between the baseline and AmE post shadowing sessions. We examined whether the extent to which shadower participants shifted their TRAP vowel was influenced by the extent of their DRESS vowel shift in the AmE post-shadowing session. If the movement of TRAP results from an attempt to maintain the phonemic intelligibility, then it should most affect the speakers who showed most significant movement of DRESS; i.e., the extent of the modification of TRAP vowels should be in parallel with the extent of the imitation of DRESS vowels. We calculated two differential values: ΔTRAP (each shadower's mean F1 value for TRAP in the AmE post-shadowing session minus that in the baseline session) and ΔDRESS (each shadower's mean F1 value of DRESS in the AmE post-shadowing session minus that in the baseline session). Greater positive values indicate stronger convergence effects, and zero means no convergence effect. **Figure 8** was constructed using the two differential values calculated for each shadower. The x axis represents ΔDRESS, the y axis represents ΔTRAP, and the red straight line indicates the reference line where $x = y$. If the TRAP shifts are proportional to the DRESS shifts, the individual data points would be expected to fall along the reference line. **Figure 8** shows that these two shifts do not necessarily occur in parallel. Many of the individual data points deviate from the reference line. The participant with the greatest convergence effect in DRESS (i.e., the rightmost data point in **Figure 8**) does not exhibit the strongest convergence effect in TRAP, and the participant with the greatest convergence effect in TRAP (i.e., the uppermost data point in the figure) does not exhibit the strongest convergence effect in DRESS. The Pearson correlation between ΔDRESS and ΔTRAP is 0.22 with a p value of .31. Taken together, these post hoc analyses suggest that the extent of the modification of TRAP vowels is not directly related to the extent of the imitation of DRESS vowels. These nonparallel shifts appear inconsistent with the intelligibility-based account.

On the other hand, the social-category-activation account makes no particular prediction about the parallelism between the extent of the modification of TRAP vowels and that of the imitation of DRESS vowels. Speakers may vary in the degree to which they have variants of these vowels explicitly indexed as American. The previous studies argue that NZE speakers are less sensitive to the dialectal differences for DRESS vowels than TRAP vowels although the two vowels in NZE are both higher than those in the other varieties of English (Hay et al., 2006; Babel, 2010; Walker & Campbell-Kibler, 2015) That is, a lower TRAP vowel may be strongly indexed as American. This may be why some shadower participants exhibited strong convergence effects in TRAP (i.e., higher ΔTRAP values) but weak or no convergence effects in DRESS (i.e., low or zero ΔDRESS values). In addition, more participants fall above the reference line in **Figure 8**, suggesting that TRAP shifts are larger than DRESS shifts in the AmE postshadowing session. In

this way, the social-category-activation account offers a coherent explanation for the asymmetry between the DRESS and TRAP shifts.

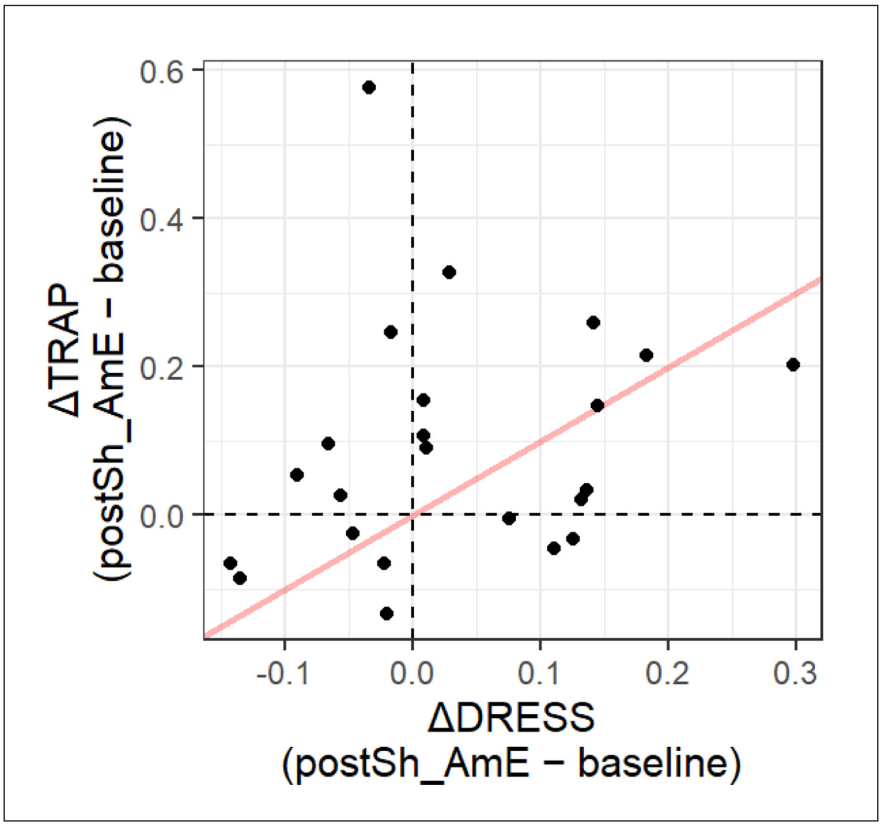

Figure 8: ΔTRAP vs. ΔDRESS with reference line where x = y.

This account relies on the assumption that NZE speakers may store exemplars encoding AmE speech, which are accessible during speech production. This view aligns well with the previous studies. Wade (2022) argues that non-Southern AmE speakers may store exemplars with Southern dialect features, and Lee-Kim & Chou (2024) suggest that Taiwan Mandarin speakers may retain exemplars encoding a different social dialect. Relatedly, Love & Walker (2013) provide the evidence that AmE speakers may store exemplars that encode British English speech. The current study provided further evidence that exposure to a different phonological system—whether regional or social—can have a broad impact on mental representations.

4.2.2. Effect of minimal pair competitor on imitation

The second key research question relates to the effect of the presence of a minimal pair competitor on the extent of phonetic imitation. In particular, we explored whether shadowers imitate model speech stimuli if imitating the model speech stimuli has a risk of blurring distinction between

minimal pair words. As discussed in Section 1, TRAP vowels in NZE are phonetically similar to DRESS vowels in AmE (Hay, 2008: 24; Ladefoged & Johnson, 2015: 234). That is, the pronunciation of *bed* in AmE is perceptually close to that of *bad* in NZE, and thus the AmE pronunciation of *bed* may be lexically ambiguous. If NZE speakers converge towards AmE speakers in their pronunciation of DRESS, the lexical contrast between *bed* and *bad* may become less clear than usual. The imitation of AmE DRESS may affect the lexical intelligibility of some DRESS words which have minimal pair competitor words with TRAP (e.g., *bed* and *guess*), whereas it may not affect the lexical intelligibility of some DRESS words which do not have minimal pair competitor words with TRAP (e.g., *deaf* and *jet*).

Our prediction was contingent on the results of research question 1. If there was no social category activation, and thus no effect on TRAP, (i.e., AmE DRESS and the local TRAP vowel are somewhat activated), then speakers may aim to minimize the overlap between minimal pair words. Under this scenario, NZE speakers may imitate model speech stimuli with AmE DRESS vowels to a smaller extent if the DRESS words have minimal pair competitors with TRAP vowels. This prediction was on the basis of the hypotheses about the intelligibility biases and the memory storage processes which can lead to minimal pair effects in speech production (see Section 1.3). On the other hand, if the social category activation affects both DRESS and TRAP (i.e., AmE DRESS and TRAP are both activated), then the contrast between them is safely maintained, and a minimal pair needn't affect pronunciation.

There was no minimal pair effect in the study. As shown in **Figure 7**, DRESS vowels were produced with slightly higher F1 values in the AmE and NZE shadowing sessions when the words have minimal pair competitor words with TRAP vowels than when they do not have minimal pair competitor words with TRAP vowels. This is in the opposite direction than predicted. However, these differences did not reach statistical significance within either of the two shadowing sessions ($ps > .05$). These results suggest that the difference between the two types of words is modulated by *task*, but they are not modulated enough to change the significance of the effects of minimal pair competitor existence for the five key sessions. It could be the case that these nonsignificant differences may reflect the fact that the two model talkers produced DRESS vowels with slightly higher F1 values in relation to the DRESS words that have minimal pair competitor words with TRAP vowels than the DRESS words that do not have minimal pair competitor words with TRAP (see **Figure 3** and **Table 2**). Recall that the difference was also nonsignificant in relation to the model speech stimuli (see Section 2.2). That is, these differences in shadowed utterances and model stimuli may be attributable to random variation rather than a systematic effect (see Sonderegger 2023: 51).

The lack of a minimal pair competitor effect on the extent of phonetic imitation is consistent with the interpretation that the participants may have accomplished a degree of "accent normalization" (Evans & Iverson, 2004), and they adjusted to AmE speech in the AmE shadowing

session. As discussed in Section 1.2, NZE speakers are generally exposed to and familiar with the American accent (Walker & Campbell-Kibler 2015; Gibson 2019). The questionnaire results indicate that the majority of the participants recognized the AmE model talker as a North American English speaker (see Section 2.3). These observations suggest that the social category "North America" was activated by the AmE model speech stimuli, making the exemplars encoding AmE speech more readily available during speech perception. That is, the spreading activation took place between the social category and the associated exemplars (see Section 1.3), with the result that more exemplars encoding AmE speech may have become more readily available. Consequently, when the participants heard the model speech stimuli with AmE DRESS vowels, the model stimuli might not have caused lexical ambiguity between the minimal pair words. As a result, the AmE model stimuli could be successfully stored as new exemplars regardless of the presence of a minimal pair competitor. Likewise, shadowers would not need to enhance the intelligibility of their shadowed utterances as a function of minimal pair competition. This interpretation suggests that even limited exposure to a different dialect can raise the activation of a social category and the associated exemplars, which may subsequently be drawn upon in speech perception. This interpretation is consistent with the discussion presented in Section 4.2.1. The key research question (2) is worth investigating while controlling for potential dialect priming effects. For example, future studies could employ acoustically modified NZE model speech instead of AmE model speech to investigate the extent to which NZE speakers imitate words with and without minimal pair competitor words. This remains an open question for future research.

Note that this interpretation of the absence of minimal pair effects supports the dual-representation model. According to Garrett and Johnson (2013, 87–88), the set of exemplars available for speech production may be different from that available for speech perception. As discussed in the preceding sections, accounting for the relatively modest shifts observed for DRESS and TRAP requires positing that a large number of exemplars encoding NZE speech continue to contribute to the creation of a production target even after exposure to AmE model speech stimuli. By contrast, as argued in this section, a large number of exemplars encoding AmE may be readily available in speech perception, such that exposure to AmE model stimuli does not induce lexical ambiguity between minimal pair words. That is, our results suggest that speech perception is more strongly influenced by spreading activation between social categories and their associated exemplars than is speech production.

Before concluding this section, it is worth mentioning that the minimal pair effects were not evident in the baseline session either. The variable *minPair* did not have significant effects on $Bark1_{DRESS}$ in the baseline session (see **Table 8**). It was also nonsignificant in relation to TRAP vowels ($p > .05$), and the interaction between *minPair* and *task* was removed in the model selection process (see Section 3.2). **Figure 9** visualizes the interaction in relation to $Bark1_{TRAP}$.

As with **Figure 7**, the y-axis represents mean *Bark1*, the *x* axis represents *task*, and the fill colour represents *minPair*. Dots represent mean values, and error bars represent standard error. These findings contradict the previous studies (Baese-Berk & Goldrick, 2009; Wedel et al. 2018; Stern & Shaw, 2023), which report that phonemes tend to be produced with more contrastive phonetic values when words have minimal pair competitors involving those phonemes in the lexicon (see Section 1.2). Why did this apparent discrepancy occur? One reason may be that the previous studies experimentally or statistically controlled for multiple lexical factors such as word frequency and segmental predictability. Our focus was on the effect of the presence of a minimal pair competitor on the extent of phonetic imitation, and we did not control for these lexical factors. Another reason may be that specific minimal pair contrasts (e.g., initial-stop voicing or neighbouring vowel phonemes) do not exert as strong an effect on hyperarticulation as anticipated. In fact, Fox et al. (2015) argue that hyperarticulation of a phoneme is conditioned by the entire phonological neighbourhood of the word (see also Wright 2004; Scarborough 2010) rather than by a specific minimal pair contrast involving the phoneme.[7]

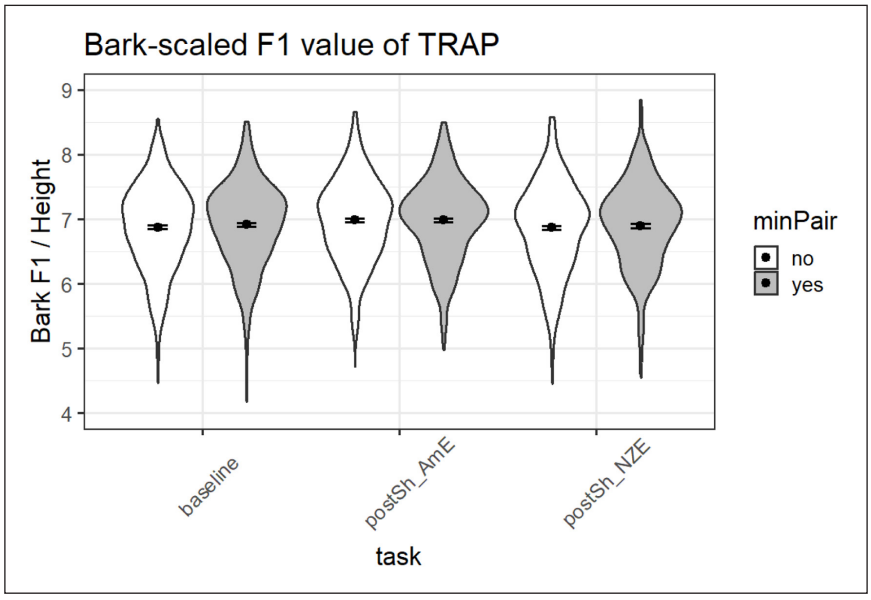

Figure 9: Violin plot of $Bark1_{TRAP}$ as a function of *task* and *minPair*. The dots show mean Bark F1 values, and error bars represent standard error.

To assess these possibilities, we conducted a post hoc analysis to explore the effects of *minPair* on baseline production while statistically controlling for multiple lexical factors (i.e., word frequency, segmental predictability, and neighbourhood density). The lexical factors were calculated using CELEX2 (Baayen et al., 1995) as with the presence of a minimal pair

[7] We deeply thank an anonymous reviewer for bringing this literature to our attention.

competitor (see Section 2.2). Word frequency information was provided in CELEX2, which was log-transformed. Segmental predictability of a target phoneme was calculated using all phonemes except the target phoneme within the same word as context and was log-transformed. Neighbourhood density was determined by counting the number of words created by inserting, deleting, or substituting a single phoneme in the given word. We fitted mixed-effects linear regression of these variables on $Bark1_{DRESS}$ and $Bark1_{TRAP}$ respectively, with respect to baseline tokens. **Table 9** shows the summary tables of these post hoc analyses. The analysis revealed that the variable *minPair* remained nonsignificant with respect to DRESS and TRAP. Amongst the lexical factors, only the variable *neighbourhoodDensity* reached significance in the two models. DRESS is produced with lower F1 values (i.e., a higher tongue position) and TRAP with higher F1 values (i.e., a lower tongue position) when the neighbourhood density of a word is greater. These effects can be interpreted as hyperarticulation because DRESS is the highest short front vowel and TRAP the lowest in the NZE vowel system (see **Figures 2** and **4**). They were produced with more contrastive formant values when the neighbourhood density is greater. Our finding is consistent with Fox et al. (2015) and may support the hypothesis that neighbouring words are globally activated in speech planning of a target word. It remains an open question under what condition specific minimal pair competitors or neighbouring words play a significant role in hyperarticulation. Addressing this question requires substantial empirical research.

Variable	Estimate	Std. Error	*t* value	Pr(>|t|)
DRESS baseline				
(Intercept)	4.8022	0.2285	21.01	< 2e–16
minPair yes	0.0277	0.0639	0.434	0.66
wordFrequency	−0.0018	0.0261	−0.071	0.94
segmentalPredictability	−0.0287	0.025	−1.149	0.25
neighbourhoodDensity	−0.0090	0.0042	−2.107	0.04
TRAP baseline				
(Intercept)	6.6817	0.1821	36.68	< 2e–16
minPair yes	−0.0524	0.0612	−0.857	0.39
wordFrequency	0.0041	0.0222	0.188	0.85
segmentalPredictability	−0.0111	0.0239	−0.468	0.64
neighbourhoodDensity	0.008	0.0037	2.141	0.03

Table 9: Summary tables of *Bark1* ∼ *minPair* + *wordFrequency* + *segmentalPredictability* + *neighbourhoodDensity* + (1 | *participant*) + (1 | *word*) for each target vowel in the baseline session. The reference level is set as *minPair* no.

5. Conclusion

In summary, this study provides further evidence that speakers converge towards a model talker with a different dialectal background when and after shadowing model speech stimuli. It was demonstrated that NZE speakers converge towards an American English (AmE) model talker in their production of DRESS and TRAP vowels after shadowing AmE DRESS vowels, even without shadowing AmE TRAP vowels. Imitation of a particular vowel from a different dialect can be systematically generalized to another vowel phoneme. This supports the idea, as previously seen in other tasks, that social category activation can affect pronunciation of phonemes, even in the absence of immediate exposure to these phonemes. We also found that minimal pair competitors with TRAP do not affect the pronunciation of DRESS vowels. The absence of this effect is consistent with the interpretation that listeners are making an overall adjustment to a US accent, shifting both vowels, such that lexical contrast between minimal pair words is maintained. Together with previous work, these findings enhance our understanding of how variants from a different dialect are stored in memory and influence postexposure speech production.

Additional files

The data and scripts used in Sections 3 are shared here: https://osf.io/nva4x/.

Acknowledgements

The earlier version of this study was presented at NZILBB Research Meeting (New Zealand Institute Language Brain and Behaviour) and LabPhon 19 Satellite Workshop: Phonetic Imitation Workshop (Hanyang University). We deeply thank the audiences for their comments and feedback at these events. We are also grateful to the editor and the anonymous reviewers for their valuable comments. The experiment was approved by Human Ethics Committee at the University of Canterbury in 2024.

Funding information

This study is supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grant number 23K12157.

Competing interests

The authors have no competing interests to declare.

Author's contributions

Daiki Hashimoto: Conceptualization, Methodology, Software, Experiment, Statistical analysis, Writing (original draft, review, and editing), Project administration, Funding acquisition. **Jen Hay:** Conceptualization, Methodology, Statistical analysis, Writing (original draft and review)

References

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Philadelphia: Linguistic Data Consortium (University Pennsylvania).

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. Berkeley: University of California, Berkeley dissertation.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5*(1), 123–150. https://doi.org/10.1515/lp-2014-0006

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554. https://doi.org/10.1080/01690960802299378

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–38. arXiv:1406.5823. https://doi.org/10.18637/jss.v067.i01

Bauer, L., & Warren, P. (2008). New Zealand English: Phonology. In K. Burridge, & B. Kortmann (Eds.) *Varieties of English: The Pacific and Australia* (pp. 39–63). Mouton de Gruyter. https://doi.org/10.1515/9783110208412.1.39

Bell, A. (1984). Language style as audience design. *Language Society*, *13*, 145–204. https://doi.org/10.1017/S004740450001037X

Bell, A. (2014). *The guidebook to sociolinguistics*. New Jersey: Wiley-Blackwell. https://doi.org/10.1002/9781394260874

Biro, T., Toscano, J. C., & Viswanathan, N. (2022). The influence of task engagement on phonetic convergence. *Speech Communication*, *138*, 50–66. https://doi.org/10.1016/j.specom.2022.02.002

Boersma, P., & Weenink, D. (2020). Praat version 6.1.14. http://www.fon.hum.uva.nl/praat/ (accessed May 28, 2020).

Cao, G. W. (2023). Phonetic dissimilarity and L2 category formation in L2 accommodation. *Language and Speech*, *67*(2), 301–345. https://doi.org/10.1177/00238309231182967

Clopper, C. G., Dossey, E., & Gonzalez, R. (2024). Raw acoustic vs. normalized phonetic convergence: Imitation of the Northern Cities Shift in the American Midwest. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *15*(1), pp. 1–34. https://doi.org/10.16995/labphon.10893

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://www.jspsych.org/6.3/about/about/

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(2–3), 145–173. https://doi.org/10.1159/000107914

Di Paolo, M., Yaeger-Dror, M., & Wassink A. B. (2010). Analyzing vowels. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 87–106). New York: Routledge.

Drager, K., & Kirtley, M. J. (2016). Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 1–24). Cambridge University Press. https://doi.org/10.1017/CBO9781139680448.003

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, *4*, 346. https://doi.org/10.3389/fpsyg.2013.00346

Eckert, P. (2016). Variation, meaning and social change. In N. Coupland (Ed.), *Sociolinguistics: Theoretical debates* (pp. 68–85). Cambridge University Press. https://doi.org/10.1017/CBO9781107449787.004

Evans, B. G., & Iverson, P. (2004). Vowel normalisation for accent: An investigation of best exemplar locations in Northern and Southern British English sentence. *Journal of Acoustical Society of America*, *115*(1), 352–361. https://doi.org/10.1121/1.1635413

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*(6), 3814–3826. https://doi.org/10.1121/1.2722209

Fiasson, R. G. M. (2015). *Allophonic imitation within and across word positions*. Christchurch: University of Canterbury dissertation.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438. https://doi.org/10.1016/j.wocn.2005.08.002

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, *83*, 97–117. https://doi.org/10.1016/j.jml.2015.04.002

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In Yu, A. (Ed.) *Origins of sound change: Approaches to phonologization*. (pp. 51–97). Cambridge: Cambridge University Press. https://doi.org/10.1093/acprof:oso/9780199573745.003.0003

German, J. S., Carlson, K., & Pierrehumbert, J. B. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics*, *41*(3–4), 228–248. https://doi.org/10.1016/j.wocn.2013.03.001

Gibson, A. (2024). Pop Song English as a supralocal norm. *Language in Society*, *53*(3), 471–498. https://doi.org/10.1017/S0047404523000131

Gibson, A., & Hay, J. (2025). Perceptual style-shifting across singing and speech: Music activates pop song English for NZ listeners. *Language and Speech*. https://doi.org/10.1177/00238309251353527

Gibson, A. M. (2019). *Sociophonetics of popular music: Insights from corpus analysis and speech perception experiments*. Christchurch: University of Canterbury dissertation.

Gnevsheva, K., Szakay, A., & Jansen, S. (2021). Phonetic convergence across dialect boundaries in first and second language speakers. *Journal of Phonetics*, *89*, 101110. https://doi.org/10.1016/j.wocn.2021.101110

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. https://psycnet.apa.org/doi/10.1037/0033-295X.105.2.251

Goldinger, S. D. (2000). The role of perceptual episodes in lexical processing. *Proceedings of Spoken Word Access Processes*, 155–159.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review, 11*(4), 716–722. https://doi.org/10.3758/BF03196625

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its origin and evolution*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486678

Hall, K., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard*, *4*, 20170027. https://doi.org/10.1515/lingvan-2017-0027

Hashimoto, D. (2019). Sociolinguistic effects on loanword phonology: Topic in speech and cultural image. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1): 11, 1–34. https://doi.org/10.5334/labphon.187

Hashimoto, D., Gomi, K. & Shiraishi, R. (2022). Updating L2 phonetic memories: Shadowing training vs. listening training. *Journal of Phonetic Society of Japan, 26*, 13–26. https://doi.org/10.24467/onseikenkyu.26.1_13

Hashimoto, D., Ida, A., & Jozuka, H. (2025). Word frequency effects on L2 learners' phonetic imitations. *JASA Express Letters*, *5*(5), 055201, 1–8. https://doi.org/10.1121/10.0036512

Hay, J., Nolan A., Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*, 351–379. https://doi.org/10.1515/TLR.2006.014

Hay, J. B., Maclagan, M., & Gordon, E. (2008). *New Zealand English*. Edinburgh University Press. https://doi.org/10.1515/9780748630882

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91. https://doi.org/10.1016/j.cognition.2015.02.012

Johnson, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499. https://doi.org/10.1016/j.wocn.2005.08.004

Johnson, K. (2012). *Acoustic and auditory phonetics*. Wiley-Blackwell.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156. https://doi.org/10.1515/labphon.2011.004

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15. https://doi.org/10.1016/j.jml.2006.07.010

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Kwon, H. 2019 The role of native phonology in spontaneous imitation: Evidence from Seoul Korean. *Laboratory Phonology: Journal of the Association for Laboratory Phonology 10*(1), 10, 1–24. https://doi.org/10.5334/labphon.83

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110167467

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics*. Cengage.

Lee-Kim, S. I., & Chou, Y. C. (2024). Unmerging the sibilant merger via phonetic imitation: Phonetic, phonological, and social factors. *Journal of Phonetics*, *103*, 101298. https://doi.org/10.1016/j.wocn.2024.101298

Lev-Ari, S. & Peperkamp, S. (2014). An experimental study of the role of social factors in language change: The case of loanword adaptations. *Laboratory Phonology*, *5*, 379–401. https://doi.org/10.1515/lp-2014-0013

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Kluwer. https://doi.org/10.1007/978-94-009-2037-8_16

Love, J., & Walker, A. (2013). Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and Speech*, *56*(4), 443–460. https://doi.org/10.1177/0023830912453132

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

MacLeod, B., & Di Lonardo Burr, S. M. (2022). Phonetic imitation of the acoustic realization of stress in Spanish: Production and perception. *Journal of Phonetics*, *92*, 101139. https://doi.org/10.1016/j.wocn.2022.101139

Martin, D. W. (2007). *Doing psychology experiments*. Wadsworth.

Mercado III, E., Mantell, J. T., & Pfordresher, P. Q. (2014). Imitating sounds: A cognitive approach to understanding vocal imitation. *Comparative Cognition & Behavior Reviews*, *9*, 1–57. https://psycnet.apa.org/doi/10.3819/ccbr.2014.90002

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422–432. https://doi.org/10.1177/026192702237958

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14. http://www.jstatsoft.org/v60/i02/

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132–142. https://doi.org/10.1016/j.wocn.2010.12.007

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*, 637–659. https://doi.org/10.3758/s13414-016-1226-0

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, *78*, 1–17. https://doi.org/10.1016/j.jml.2014.10.003

Pierrehumbert, J. B. (2001). Exemplar dynamics, word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 135–157). Benjamins. https://doi.org/10.1075/tsl.45.08pie

R Core Team. (2024). R version 4.3.3. https://www.r-project.org/.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language awareness*, *20*(1), 45–59. https://doi.org/10.1080/09658416.2010.540326

Sanchez, K., Hay, J., & Nilson, E. (2015). Contextual activation of Australia can affect New Zealanders' vowel productions. *Journal of Phonetics*, *48*, 76–95. https://doi.org/10.1016/j.wocn.2014.10.004

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory Phonology*, *10*, 557–586. https://doi.org/10.1515/9783110224917.5.557

Shockley, K., Sabadini, L. & Fowler, C. A. (2004). Imitation in shadowing words. *Perception and Psychophysics, 66*(3), 422–429. https://doi.org/10.3758/BF03194890

Sonderegger, M. (2023). *Regression modeling for linguistic data*. MIT Press.

Sóskuthy, M. (2015). Understanding change through stability: A computational study of sound change actuation. *Lingua, 163*, 40–60. https://doi.org/10.1016/j.lingua.2015.05.010

Stern, M. C., & Shaw, J. A. (2023). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language*, *132*, 104443. https://doi.org/10.1016/j.jml.2023.104443

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in psychology*, *4*, 1015. https://doi.org/10.3389/fpsyg.2013.01015

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, *37*(3), 276–296. https://doi.org/10.1016/j.wocn.2009.03.004

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, *185*, 1–20. https://doi.org/10.1016/j.cognition.2019.01.004

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97–100. https://doi.org/10.1121/1.399849

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329. https://doi.org/10.1111/1467-9280.00064

Wade, L. (2022). Experimental evidence for expectation-driven linguistic convergence. *Language*, *98*, 63–97. https://doi.org/10.1353/lan.2021.0086

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, *6*, 546. https://doi.org/10.3389/fpsyg.2015.00546

Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*, 247–274. https://doi.org/10.1515/TLR.2006.010

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61–88. https://doi.org/10.1016/j.jml.2018.01.001

Wells, J. C. (1982). *Accents of English*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611759

Wright, R. (2004). Factors of lexical competition in vowel articulation. *Laboratory Phonology*, *6*, 75–87.

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS one*, *8*(9), e74746. https://doi.org/10.1371/journal.pone.0074746