Journal of Phonetics 88 (2021) 101076

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Phonetic convergence to non-native speech: Acoustic and perceptual evidence

Mónica A. Wagner [a,b,*], Mirjam Broersma [b], James M. McQueen [a,c], Sara Dhaene [a], Kristin Lemhöfer [a]

[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[b] Centre for Language Studies, Radboud University, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands
[c] Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 25 February 2020
Received in revised form 3 June 2021
Accepted 10 June 2021

Keywords:
Vocal accommodation
Phonetic convergence
Non-native speech
Shadowing
Acoustic analysis
AXB task

ABSTRACT

While the tendency of speakers to align their speech to that of others acoustic-phonetically has been widely studied among native speakers, very few studies have examined whether natives phonetically converge to non-native speakers. Here we measured native Dutch speakers' convergence to a non-native speaker with an unfamiliar accent in a novel non-interactive task. Furthermore, we assessed the role of participants' perceptions of the non-native accent in their tendency to converge. In addition to a perceptual measure (AXB ratings), we examined convergence on different acoustic dimensions (e.g., vowel spectra, fricative CoG, speech rate, overall f0) to determine what dimensions, if any, speakers converge to. We further combined these two types of measures to discover what dimensions weighed in raters' judgments of convergence. The results reveal overall convergence to our non-native speaker, as indexed by both perceptual and acoustic measures. However, the ratings suggest the stronger participants rated the non-native accent to be, the less likely they were to converge. Our findings add to the growing body of evidence that natives can phonetically converge to non-native speech, even without any apparent socio-communicative motivation to do so. We argue that our results are hard to integrate with a purely social view of convergence.

1. Introduction

The tendency of speakers to align their speech to that of another speaker has proven to be subtle and highly variable. This phenomenon, known as *speech accommodation*, has been demonstrated across different experimental settings (e.g., spontaneous interactions and both interactive and non-interactive tasks in the laboratory) and at different levels of speech processing (e.g., syntactic, lexical, acoustic–phonetic). When alignment takes place on the acoustic–phonetic level, it is more specifically referred to as *vocal accommodation* or *phonetic convergence* (Natale, 1975; Pardo, 2006). Over the last decades, phonetic convergence has gained prominence in communication theories (Pardo, Urmanche, Wilman, & Wiener, 2017), even being proposed as a vehicle of sound change (e.g., Delvaux & Soquet, 2007; Giles, Coupland, & Coupland, 1991; Pardo, 2016; Trudgill, 2008). However, most studies to date have analyzed phonetic convergence between native speakers of a language, while today, in most parts of the world, conversations in which at least one person is speaking in a second language (L2) are commonplace (Costa, Pickering, & Sorace, 2008). In this article, we ask whether convergence can occur not only between native speakers of a language, but also to a variety that natives are not generally inclined to adopt, namely non-native speech. Moreover, we test whether such convergence can occur in a non-interactive setting, where socio-communicative motivations are minimized. Together, these conditions allow us to examine the basic processes involved in phonetic convergence.

1.1. Theoretical accounts of speech accommodation

Despite its prevalent nature, little is known about the mechanisms underlying accommodation generally and phonetic convergence specifically, with various theories having been developed to account for why these phenomena occur. On the one hand, accounts such as that put forth by Communication Accommodation Theory (CAT; Gasiorek, Giles, & Soliz,

* Corresponding author at: Donders Institute for Brain, Cognition and Behaviour, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.
E-mail address: monica.wagner@donders.ru.nl (M.A. Wagner).

https://doi.org/10.1016/j.wocn.2021.101076
0095-4470/

2015; Giles et al., 1991; Shepard, Giles, & Le Poire, 2001) attribute a social function to alignment. According to the CAT, accommodation serves to decrease, maintain, or increase (in the case of *divergence*) social distance, usually to garner approval (Giles & Ogay, 2007). Studies showing that phonetic convergence can be influenced by social factors such as attitude towards the speaker (Babel, 2009, 2010; Yu, Abrego-Collier, & Sonderegger, 2013) or the speaker's role in an interaction (Giles, 1973; Pardo, 2006) lend support to such accounts.

On the other hand, the observation of accommodation in non-interactive settings has been taken as evidence of the involvement of low-level cognitive processes, essentially, that "humans are hardwired to imitate" (Coles-Harris, 2017, p. 14). Such is the case for another set of accounts maintaining that accommodation is an automatic result of the tight link between perception and production, although varying in the precise mechanism proposed. For example, while for Goldinger, (1998) and (Goldinger and Azuma, 2004) convergence results from the heightened detail of episodic memory traces which are evoked during production, the motor theory of speech perception (Liberman & Mattingly, 1985) and direct realist theory (Fowler, 1986; Fowler, Brown, Sabadini, & Weihing, 2003; Sancier & Fowler, 1997) posit direct perception of articulatory gestures which are primed during production. Pickering and Garrod (2004), in their interactive alignment account, further argue that the shared perception-production representations serve the purpose of aligning interlocutors' representations in order to increase mutual understanding.

The fact that there is evidence that both social and perceptual-motor factors are at play during accommodation reveals that these two overarching accounts are not mutually exclusive (Coles-Harris, 2017), leading to the formulation of hybrid views (Babel, 2012; Walker & Campbell-Kibler, 2015). For example, some have argued that phonetic convergence may be an automatic process that can be inhibited or facilitated by social factors (Babel, 2012; Dijksterhuis & Bargh, 2001). Similarly, Pickering and Garrod (2013) have incorporated social effects into their mechanistic account of accommodation, putting forth the possibility of both automatic and non-automatic processes of imitation. Gambi and Pickering (2013) even speculate that listeners may recur less to automatic imitation with speakers they perceive as dissimilar to themselves, resulting in less convergence. The latter is particularly relevant in the case at hand: convergence to non-native speakers.

Since so far studies on phonetic convergence have almost exclusively focused on convergence between native speakers, it remains to be determined how much of what is known about phonetic convergence can be generalized to communication involving non-native speakers. Studying convergence to non-native speech can therefore extend our understanding to a case that is underrepresented in the scientific literature given its occurrence (i.e., native-non-native speaker interaction). Moreover, Lewandowski and Nygaard (2018) have argued that studying convergence to non-native speech can shed light on the issue of its underlying mechanisms and help disentangle the role of different factors. More generally, determining whether native speakers of a language tend to adopt phonetic patterns from someone who learned the language as a second language offers an interesting and strong test of the flexibility of speech.

1.2. Eliciting phonetic convergence

The results of any study on convergence, and hence any interpretations that can be derived therefrom, hinge crucially on the task used to elicit convergence. The tasks most frequently recurred to in the lab are conversational interaction, perceptual exposure, and shadowing. Spontaneous conversation is usually elicited in the lab via a task participants have to resolve together, such as spotting the difference between pairs of images (e.g., the Diapix task; Van Engen et al., 2010) or maps (e.g., the HCRC Map Task, Anderson et al., 1991; Brown, Anderson, Yule, & Shillcock, 1983; the MMT, Pardo et al., 2019). In contrast, during perceptual exposure tasks, participants are asked to just passively listen to speech (e.g., Delvaux & Soquet, 2007; Yu et al., 2013) or perform a task such as identifying previously heard items (e.g., Kim, 2012). In so-called shadowing tasks (i.e., auditory naming; e.g., Goldinger, 1998; Goldinger & Azuma, 2004; see Pardo et al., 2017 for a review), participants listen to and repeat the speech of a model speaker, usually single words. All of these tasks allow researchers to ensure participants are exposed to and produce the target sounds. Convergence is then usually measured by comparing participants' baseline speech from early on in the task (or before exposure to the input) to that from later in the task (or after exposure).

All of these tasks have different strengths and a researcher may opt for one or another based on their research questions and priorities. As Felker, Troncoso-Ruiz, Ernestus, and Broersma (2018)[1] point out, face-to-face interactive tasks have high ecological validity, i.e. they are highly representative of real-life situations. In addition, the distraction of resolving a task together typically leaves the participants naïve with respect to the actual purpose of the study (i.e., measuring the degree of phonetic convergence). Thus, lab-induced behaviors or strategies by the participant, such as nervousness or attempts to act according to the experimenter's (supposed) expectations, are minimized in interactive tasks.

In contrast, non-interactive settings such as auditory naming and perceptual exposure have the advantage of tighter experimental control over the exact quality, amount, and timing of phonetic exposures (Felker et al., 2018). While confederates themselves will usually vary in how they pronounce an item (e.g., Broos et al., 2016; Rao, 2013), or conceivably even converge to or diverge from the participants, this variation is absent when the input is pre-recorded rather than uttered in real-time (Felker et al., 2018). Furthermore, socio-motivational factors such as common communicative goals (Pardo et al., 2017) and strategies (Lewandowski & Nygaard, 2018) or more psycho-social effects such as the perceived status and attractiveness (Babel, 2012) of the interlocutor are all potential modulators of convergence (Babel, 2009, 2010) that

[1] Note that the "ventriloquist paradigm" introduced in that study, in which participants interact face-to-face with a confederate whose vocal responses are actually pre-recorded, could allow for high ecological validity while maintaining experimental control. However, here we did not opt for this paradigm in order to avoid socio-motivational effects due to the physical presence of and interaction with another person, which would likely vary across participants.

are minimized here. Thus, any convergence elicited in a non-interactive setting may be less subject to the influence of socio-motivational factors. It is, however, important to note that social factors can and have been found to still be at play in non-interactive tasks (e.g., attitudes to the speaker's dialect group or voice; Abrego-Collier, Grove, Sonderegger, & Yu, 2011; Babel, 2009, 2010; Babel, McGuire, Walters, & Nicholls, 2014), although presumably to a smaller degree.

In the present study, we opted for a non-interactive repetition task in order to maximize experimental control over the stimuli as well as minimize the potential influence of social factors, particularly those resulting from in-person experience with a confederate and communicative goals. This was to avoid the incidence of social factors that could boost convergence or introduce additional between-subject variation. In order to also keep participants naïve about the purpose of the study, we disguised the task as a memory game where word repetition appeared as a natural component.

1.3. Measuring phonetic convergence

Phonetic convergence is usually measured acoustically or perceptually. Perceptual measures consist in judgments from an AXB perceptual similarity task in which a separate set of raters hears the participants' baseline (A) and post-exposure (B) utterances and judges which of the two sounds more like the target item (X) the participant was exposed to (Goldinger, 1998). Perceptual measures are generally considered more holistic measures (Aguilar et al., 2016; Pardo, Urmanche, Gash, et al., 2018). A drawback is that they do not provide much information about what exactly speakers are converging to. Acoustic measures, which evaluate the degree of a participant's change in a certain acoustic dimension by comparing participants' baseline and post-exposure distances to the target speaker, can be more informative in this respect. However, previous studies have found varying patterns of convergence across different sounds and acoustic dimensions (e.g., Babel, 2009; Levitan & Hirschberg, 2011; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013; Walker & Campbell-Kibler, 2015), with participants even converging on one dimension while diverging on another. Furthermore, it is impossible to measure everything and participants may indeed converge to target speech along a dimension not measured (Pardo et al., 2013). Although ratings, in which these acoustic patterns are combined into a single precept, may provide a simple solution, discovering the sources of acoustic variability can also prove informative to theories of convergence. For example, Babel (2009) found that native English speakers converged most to low vowels, which she attributed to greater production variability for these sounds (also: Walker & Campbell-Kibler, 2015, but see Pardo, 2010). Moreover, so far little is known about how different acoustic features come together in raters' perception, as well as what acoustic dimensions might be most relevant to raters when judging convergence (Goldinger, 1998; Pardo et al., 2013), questions which require both types of measures. In light of the variability in convergence to different acoustic dimensions and how little is understood about how they combine into one percept, Pardo and collaborators (2013, 2017) have advocated the inclusion of multiple acoustic measures in addition to holistic perceptual measures, as well as analyzing the relationship between the two types of measures. Here, we have adopted this two-fold approach, including both acoustic and perceptual measures of convergence, as well as analyzing what acoustic dimensions predict the ratings. Additionally, considering how few studies have evaluated convergence to non-native speech and since we are not aware of any including the particular language combination used here, we chose to measure various different sounds and acoustic dimensions thereof to provide a more comprehensive look at the multi-faceted nature of non-native speech.

1.4. Phonetic convergence to non-native speech

Non-native speech is often associated with more negative attitudes than native speech, with native speakers tending to rate non-native speakers as having a lower social status and being less competent than native speakers (Lewandowski & Nygaard, 2018; Nelson, Signorella, & Botti, 2016). Thus, following social accounts of phonetic convergence, in the absence of communicative motivations, native speakers might generally be less inclined to sound similar to non-native speakers, resulting in less convergence and perhaps even divergence. Accordingly, in a non-interactive setting, such a modulation of the degree of convergence by the nativeness of the interlocutor would be evidence against a purely automatic mechanism for convergence. Some researchers have indeed predicted that non-native speech may be perceived as too distant or different by native speakers or that non-native status may block the occurrence of convergence (Gambi & Pickering, 2013; Walker & Campbell-Kibler, 2015). On the other hand, the potential observation that native speakers do converge to non-native speech in the absence of an interlocutor, regardless of accent and despite negative attitudes against non-native speech, would be hard to integrate with a purely social account of convergence (note, however, that social accounts would not preclude convergence to non-native speakers in interactive settings).

From an automatist, cognitive perspective, there are reasons to assume that native speakers might even converge *more* to non-natives than to fellow natives. First of all, in order for phonetic convergence to occur, some initial linguistic distance is required; that is, speakers cannot align if they are already aligned (Kim, 2012). Some researchers have therefore argued that the larger the phonetic distance, the greater the amount of convergence, simply because there is more phonetic room for it (e.g., Babel, 2012). Walker and Campbell-Kibler (2015) provided support for this claim in their auditory naming study with model speech from New Zealand, Australia, U.S. Midland, and U.S. Inland North in which they found participants to converge more to dialects farther from their own.

Yet another reason to expect more convergence to non-native versus native speech derives from what Bell (1984) refers to as "audience design." This concept encompasses all sorts of actions a speaker may take in order to respond to their interlocutor's perceived conversational demands. Some researchers have argued that, when interacting with non-natives, approximating non-native speech may serve to increase the chances of being understood (Costa et al., 2008; Kim, 2012). Native speakers do routinely adapt their speech when communicating with non-natives to make it more

understandable, as attested by "foreigner" or "clear talk" (see Wooldridge, 2001). Therefore, it is not too far-fetched to assume that natives may also adopt characteristics of their non-native interlocutor's speech (see, for example, Ivanova, Costa, Pickering, & Branigan, 2007 cited in Costa et al., 2008). However, convergence motivated on the grounds of audience design should be minimal in non-interactive tasks. This is consistent with recent evidence from a syntactic alignment study that showed that native speakers aligned to non-native but not native interlocutors, only in a condition in which the confederates did not demonstrate flexibility (indexed in that study by alternating between different syntactical structures; Hwang, Brennan, & Huffman, 2015). Thus, natives seem willing to adopt non-nativelike speech patterns but only when they think their interlocutor needs it, supporting the idea of comprehension-motivated convergence to non-native speech.

The present study aims to contribute to the discussion about whether natives converge to non-native speech. We used a novel non-interactive task that minimized the influence of socio-motivational factors and disguised the task's purpose, limiting any potential boost convergence could receive due to interaction and the need to communicate with someone with another L1. The latter is of particular relevance here given the salient nature of non-native speech, which could elicit extra attention and thus potentially the use of explicit strategies (Walker & Campbell-Kibler, 2015). These circumstances were expected to provide a strong test of whether native speakers can converge to non-native speech.

1.5. Empirical evidence of phonetic convergence to non-native speech

Empirical investigations into convergence to non-native speech have been scarce and so far the results have been inconsistent. In particular, the first two studies on phonetic convergence to non-native speech seem to present contradictory findings: while Kim, Horton, and Bradlow (2011) found that natives did not converge to non-natives, a follow-up study by the same authors (Kim, 2012) did find evidence of convergence. These differences may be due to methodological differences.

The first study, Kim et al. (2011), made use of a Diapix experiment in which pairs of participants, each with their own picture, had to perform a spot-the-difference task by working together without being able to see each other or the other's picture. The speakers were native American English speakers and Chinese and Korean L2 speakers of English who participated in pairs of either the same L1/same dialect, same L1/different dialect, or different L1 . Convergence was measured with native speaker perceptual ratings of samples of speech from early and late in the conversations. The results showed convergence only in the same L1/same dialect conversations. The authors interpreted this finding as evidence for greater convergence to more similar speakers (see, however, Walker & Campbell-Kibler, 2015 mentioned earlier).

However, Kim et al. (2011) also speculated that the low proficiency of the non-native speakers may have contributed to the lack of convergence, possibly leading to a conscious decision by the native speakers to not converge or to produce clear speech instead. In a follow-up study, Kim (2012) had native English speakers listen to recordings of model speakers reading words and sentences. For one group of participants, the model speakers were high-proficiency Korean L2 speakers of English, whereas for another group, as in Kim et al. (2011), they were native English speakers with the same or different dialect as the participants. The participants' task was to select the item they had heard from a series of items on the screen. Pre- and post-exposure productions of participants reading the items out loud were obtained and convergence was measured acoustically as well as perceptually (for sentences). Morover, the Implicit Association Task (IAT; Greenwald, McGhee, & Schwartz, 1998) was administered in order to assess participants' general attitude to foreigners. This time, the researchers observed convergence to all groups of model speakers. Moreover, they observed more convergence when an item's initial acoustic distance was larger, lending support to the involvement of low-level cognitive processes. However, while participants' attitudes towards foreigners did not directly relate to their degree of convergence, it interacted with baseline acoustic distance, with its effect varying depending on the acoustic dimension. The authors attributed the lack of convergence in the first study to the greater influence of psycho-social factors which may have been at play during the interactive task, as well as the low proficiency of the L2 speakers, which would have been a greater obstacle in the task with spontaneous conversation and perhaps led to other strategies (e.g., clear speech).

Since the publication of these first studies, a handful of other researchers have reported evidence of convergence to non-native speech (e.g., Broos et al., 2016; Lewandowski & Nygaard, 2018; Rao, 2013; Weise, Levitan, Hirschberg, & Levitan, 2019). However, we are aware of only one study besides Kim (2012) that has examined convergence to non-native speech with a non-interactive task, where psycho-social and socio-communicative motivations are reduced, namely the recent study of Lewandowski and Nygaard (2018). In that study, native American English speakers listened to recordings of two other native American English speakers and two Mexican Spanish L2 speakers English. Participants first performed a perceptual exposure task in which they had to identify the items heard in order to become familiar with the model speakers' voice and speech patterns, followed by auditory naming. A separate set of participants rated each item's intelligibility and accentedness, confirming that the non-native speakers were perceived as less intelligible and more accented. Furthermore, measures of listeners' attitudes to the speakers obtained with another set of participants confirmed that the non-native speakers were perceived as being of lower social status. Finally, the authors investigated whether the participants' baseline variability (e.g., vowel dispersion) in the selected acoustic measurements predicted their degree of convergence. Both acoustic and perceptual measures of convergence were employed. The rating results revealed convergence to the non-native model speakers — in fact, more than to the native model speakers, despite the non-natives' lower perceived social status. According to the authors, these patterns suggest that intelligibility and accentedness play a greater role than social attitudes. They interpret this as support for a perceptual mechanism for convergence in non-interactive tasks. However, they also suggest that the block of perceptual exposure may have increased perceptual fluency and thus

reduced any potential negative effect of social attitudes towards the non-native speech, consequently facilitating convergence to non-native speech in their study (see also Gambi & Pickering, 2013).

In the current study, we aim to extend the so-far small body of evidence of convergence to non-native speech in non-interactive settings and shed more light on the relation with accentedness, comprehensibility and familiarity. Following Lewandowski and Nygaard (2018) we incorporated data collected on how the non-native speaker's accent was perceived in terms of accentedness, comprehensibility, and familiarity, but now crucially using data obtained from the speakers themselves, to see whether this influenced the degree to which they adopted the non-native speaker's phonetic patterns.

It is important to note that in all of the studies finding convergence to non-native speech, including Kim's (2012) non-interactive study, multiple acoustic dimensions were assessed for convergence, and convergence was inconsistently observed across them. Therefore, it can also be argued that any potential convergence elicited in Kim et al. (2011), which only employed AXB ratings and did not find convergence to non-native speech, might have been too subtle or irrelevant for the raters in that study to pick up on and use in their evaluations. Following current trends in the field, the present study includes both perceptual and acoustic measures. Knowing what exactly speakers converge to can be especially informative when evaluating convergence to non-native speech because natives may converge more or less to dimensions that vary greatly from their native realizations. If more convergence is observed for distant non-native realizations, this would fit with the claim of greater convergence for larger phonetic distances. If, however, participants selectively avoid the most distinctively non-native sounds or dimensions, that could potentially indicate a limit to any "automatic" mechanism (e.g., see Walker & Campbell-Kibler, 2015).

1.6. The present study

The goal of the present study was to determine whether native speakers converge to the speech of a non-native speaker in a non-interactive setting in the absence of an interlocutor. If so, this would contribute to the growing evidence that convergence can occur to a variety that speakers are not generally inclined to adopt, and that they do so in a socially impoverished task without any apparent communicative motivation. In addition to using a holistic perceptual measure of convergence, we aimed to discover what exactly speakers converge to (or not) in this setting by measuring convergence to a variety of sounds and acoustic dimensions. Furthermore, we sought to discover what, if any, acoustic dimensions contribute to the perception of convergence in this context. In order to address these research questions, we developed a novel task where a repetition task is disguised as a memory task in which participants repeat series of words after hearing them spoken by a non-native speaker. By using this experimental task, which allows for stimulus control and minimizes the influences of socio-motivational factors, we target the question of whether natives can demonstrate phonetic convergence to non-native speech. Given that some researchers have suggested that familiarity with accent may influence tendency to converge to non-native speech (Gambi & Pickering, 2013; Lewandowski & Nygaard, 2018), here we opted for a variety that we expected to be unfamiliar to our Dutch participants: Serbo-Croatian. We further explored the role, if any, that native speakers' perception of the non-native speaker's accent plays in the tendency to converge to the model speech.

2. Method

Phonetic convergence was elicited by a repetition task disguised as a memory task and assessed in two ways: acoustically and perceptually. The acoustic analyses provided an objective measure of how much more similar (or dissimilar) the participants' speech was to the model speaker's during the convergence task relative to their baseline values from before the task on specific acoustic dimensions. The perceptual analyses were conducted on ratings obtained with a separate set of participants. These raters performed an AXB similarity judgment task in which they heard participants' speech from before and during the convergence task and had to choose which productions sounded more similar to those of the model speaker. Following the repetition task, participants provided ratings of how accented, comprehensible, and familiar they found the model speaker's speech, which were included in the analyses to see how they affected degree of observed convergence.

Studies on phonetic convergence generally tend to find that convergence increases over the span of a task (e.g., Pardo, 2006; but see Babel, 2012). Additionally, Lewandowski and Nygaard (2018) suggest that, in the case of non-native speech, increased exposure may allow for greater convergence by increasing perceptual fluency. Here we check this by seeing if there is a difference in convergence between participants' first and last repetitions of model utterances.

2.1. Participants

Female monolingually-raised native Dutch speakers were recruited for the convergence task via the Radboud University participant database and through advertisements placed locally and on social media. Only females were included because, while recent studies have found that sex may interact with different factors during convergence (Pardo et al., 2017; Weise et al., 2019), it was not the purpose of this study to test this. A total of 93 participants completed the convergence task. Because of differences in regional varieties of Dutch, only participants from two neighboring provinces (Gelderland and Noord-Brabant) were accepted in order to reduce differences in baseline distance to the model speaker due to regional differences (also see Vallabha & Tuller, 2004 who mention the possibility of a role for dialect differences in imitation bias). Twelve participants were excluded because they either self-reported another non-standard accent (six), grew up bilingually (one), or did not complete the language background questionnaire (five). An additional five participants who reportedly believed the model speaker could be a native speaker of Dutch were also removed from the analyses, although including them did not alter the results. This left a final sample of 76 participants between the ages of 18 and 28 (M = 22.2, SD = 2.5).

No participants reported any current hearing or speech problems.

A separate group of monolingually-raised native Dutch speakers was recruited with the same database to perform the AXB perceptual similarity task.[2] Here, the sex and regional requirements did not apply. No raters reported any hearing problems nor familiarity with Croatian or Croatian-accented Dutch. One rater had to be discarded due to technical error and was replaced for a final sample of 16 raters (12 female) between the ages of 18 and 29 (M = 22.8, SD = 3.4). This number satisfies the standard of 5–30 raters per token (e.g., Pardo et al., 2017), with the added benefit that here all raters rated all participants.

Ethical approval for the study was obtained from the Ethics Committee of the Faculty of Social Sciences, Radboud University, The Netherlands (ECSW-2016-1403-391 and ECSW-2018-035). Participants provided written informed consent (see also Sub-section 2.3.2). The experiment was conducted according to the ethical standards of the Declaration of Helsinki. Participants and raters received course credit or gift vouchers (€15 or €20, respectively) for their participation.

2.2. Design

The present study assessed phonetic convergence to non-native speech using single words. Four critical items were selected which each contained a sound (two vowels, two consonants) which the model speaker produced in a way that was not standard Dutch (later confirmed by participants' baseline measurements). For both the acoustic and perceptual measures of convergence, productions of the critical items from before as well as during the repetition task were necessary. Participants' baseline realizations of the critical items were obtained before the repetition task with a word reading task. During the repetition task, participants repeated the critical items after the model speaker five times each within the context of a memory task with filler items (see "Procedure" below). Some of these filler items, as well as items included for reference to the critical sounds and vowel normalization (see Sub-section 2.3.3 for more information), were also included in the word reading task. The AXB perceptual similarity task included the model speaker's utterance of one of the critical items and one of the participant's recordings from the repetition task along with their corresponding baseline production from the word reading task.

The study reported here is part of a larger individual differences study for which other tasks were administered. Because of this, the order of the stimuli was kept constant for all participants. This is also the reason why the language background questions were only able to be administered a day later, after the rest of the tasks had been completed. However, the first tasks the participants completed were the word reading task, followed by the repetition task.

2.3. Materials

2.3.1. Model speaker

The model speaker was a 32-year-old female native speaker of Serbo-Croatian from Zagreb, Croatia and self-identified her variety of Croatian as standard Shtokavian. She had been living in The Netherlands for about 10 years, had been speaking Dutch for nearly three, and rated her own Dutch proficiency a 3 on a scale from 1 (beginner) to 5 (mother tongue). Before moving to The Netherlands, the speaker had also spent several years in Germany and the Czech Republic and she was fluent in English.

The participants rated the model speaker's speech in terms of accentedness, comprehensibility, and familiarity with her accent. They rated (median values) her accentedness a moderate 5 on a scale from 1 (no foreign accent/nativelike) to 9 (very strong foreign accent), her comprehensibility a 2 (fairly comprehensible) on a scale from 1 (very comprehensible) to 9 (very incomprehensible), and a 2 (rather unfamiliar/I hear it less than once a week) on a scale from 1 (completely unfamiliar/I've never heard it before) to 4 (very familiar/I hear it a couple times a week). No participants reported regular exposure to Croatian or Croatian-accented Dutch, with the most frequent exposure reported being less than once a week (2; N = 8 for Croatian and N = 5 for Croatian-accented Dutch). Moreover, no participants were able to accurately identify the model speaker's native language, with the majority guessing German (N = 24), followed by Turkish (N = 13) and Spanish (N = 13). When asked how sure they were of these guesses on a scale from 1 (completely uncertain) to 5 (very certain), participants rated their certainty (median) 3, 2, and 2 for the three guessed languages respectively.

2.3.2. Critical sound selection

Prior to the experiment, the model speaker was asked to read aloud sentences from a set developed to assess non-native Dutch speech (Cucchiarini, Strik, & Boves, 2000). From those utterances, we chose the following four Dutch sounds for use in the critical items because the speaker realized them in a distinctively non-standard Dutch way: [t], [oː], [aː], and [s] (specifically before another consonant). Although regional differences exist which deviate from the standard realizations described below (see footnotes 3-6) and in some cases may approximate those of the model speaker, such realizations are not known to characterize our participants' regional varieties nor do we have any reason to assume familiarity with them.

The final [t] was selected because, unlike in standard Dutch, the model speaker tended to realize them with aspiration. Unlike English, in Dutch the main cue distinguishing stop voicing perceptually is not aspiration but the presence of voicing during the closure (e.g., van Alphen & Smits, 2004), with voiced stops presenting prevoicing or negative voice-onset times (VOTs) and voiceless stops realized with short-lag VOTs.[3] Thus, the model speaker's aspirated voiceless stops were saliently non-native.

The long [oː] vowel was selected because the model speaker tended to produce it with a shorter duration and as a monophthong, hence realizing it similar to the Dutch vowel [ɔ]. The mid back rounded Dutch vowel [oː] is longer than its

[2] Although, as a reviewer notes, Croatian-Dutch balanced bilingual speakers would have been preferable, this was not feasible within the limits of this study.

[3] Voiceless stops are slightly aspirated in some Eastern and Northern varieties of Dutch (e.g., Overijssel and Groningen; Collins & Mees, 2003; Goblirsch, 2015; Gussenhoven & Broeders, 1997).

counterpart [ɔ], and is even diphthongized in Northern Standard Dutch (Booij, 1999).

The [aː] was selected because the model speaker did not distinguish that vowel from another Dutch vowel, conflating the two Dutch open back unrounded vowels [aː] and [ɑ], which differ in length and backness.[4] She produced the [aː] more like a long [ɑ].

The sound [s] before [t] was selected because, unlike in standard Dutch, the model speaker sometimes produced [s] in this position in a more retracted way, like [ʃ].[5] The Dutch phonetic inventory contains the voiceless alveolar fricative [s][6] and, marginally, the postalveolar [ʃ], which only occurs in loan words and with assimilation before 'j' (Booij, 1999; Collins & Mees, 2003).

2.3.3. Items

Critical items. The following disyllabic nouns containing the critical sounds (final [t], [oː], [aː] and initial [s] before [t]) were selected for use as the critical items in the convergence task: *atleet* ("athlete"), *saldo* ("balance"), *schaduw* ("shadow"), and *stempel* ("stamp"). Given Goldinger's (1998) findings that lexical frequency may modulate phonetic convergence (but see Pardo et al., 2013, 2017), and considering that it was not the purpose of the present study to evaluate such effects, all words were low frequency (<30 per million) according to SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010). The critical sounds did not occur in any other of the items. In addition, no other items contained voiceless stops.

As mentioned in the Introduction, phonetic convergence can vary greatly across different acoustic dimensions and the source of these patterns has yet to be uncovered. Because of this, and in order to provide a more comprehensive look at convergence to the multi-faceted nature that is non-native accent, a large set of items was sacrificed here for the sake of diversity in phonetic features and acoustic dimensions.

Filler items. Fifty nouns were chosen for use as filler items. In order to distract from the purpose of the study, many of the words that were presented in a trial together were semantically related (e.g., *nose, arm, shoulder*). Nine of the filler words were used as practice items.

Reference and vowel space items. In addition to the items for the convergence task, reference items containing the closest equivalents to the non-native's realization of the critical sounds (i.e., *haardos* [ɔ] "hairstyle" for *saldo*, *schatkist* [ɑ] "treasure chest" for *schaduw*, and *crasht* [ʃ] "crashes" for *stempel*) were also selected for the participant to read out loud during the word reading task. These words were matched to the critical items in terms of phonetic context and were obtained for comparison with the repetitions of the critical items (e.g., in vowel plots). The following 12 words targeting the Dutch vowels (Adank, van Hout, & Smits, 2004) were also included in order to be able to normalize the spectra of each participant's utterances for their vowel space: *taak* [aː], *tak* [ɑ], *pet* [ɛ], *piek* [i], *kip* [ɪ], *pot* [ɔ], *koets* [u], *fuut* [y], *put* [ʏ], *keet* [eː], *poot* [oː], and *peuk* [øː]. In the end, *keet*, *poot*, and *peuk,* which are diphthongized in Northern Standard Dutch, were not used for normalization (Booij, 1999).

2.4. Procedure

2.4.1. Model utterances

The model speaker was recorded digitally at 44.1 kHz in a soundproof booth with a Shure SM57-LCE microphone placed in front of her. She read lists of words containing the critical items, filler items, and items to normalize for her vowel space for comparison to the participants (at least two utterances per sound). The words were segmented from the audio files and one token each of the critical items was selected for use in the convergence task. The tokens to be used for the filler items were also chosen at that time. Filler items for the 2–3 word sequences were concatenated and all recordings were amplitude-normalized in Audacity.

2.4.2. Convergence task

In order to prevent the use of any conscious strategies, participants were misled to believe that the purpose of the study was to test their auditory memory. They provided written informed consent prior to the experiment and again after debriefing.

Participants performed the experiment individually in a sound-attenuated recording booth. They were seated in front of a computer screen where the instructions were presented. The experiment was conducted using the program Presentation (Neurobehavioral Systems Inc., Berkeley, CA, USA). During the convergence task, audios of the model speaker were played over Sennheiser headphones at the same comfortable volume for all participants. Participants' utterances were recorded digitally at 48 kHz with a Shure SM57-LCE microphone.

Before the repetition task, participants provided baseline measures by reading aloud lists of words presented on the screen. Participants were instructed to read the words as naturally and clearly as possible. These lists contained the critical items, reference items, vowel space items, some of the filler items, as well as words for the other tasks in the larger study. Each word was presented twice to increase the chances of obtaining useable recordings.

The convergence task was disguised as a memory task. Participants listened to word sequences of varying length (1–3 words) which they were asked to immediately repeat backwards (e.g., *nose*, *arm*, *shoulder* → *shoulder*, *arm*, *nose*). Crucially, critical items were always presented as single-item trials as to minimize the influence of memory load and co-articulatory effects. As an extra control, critical trials were also always preceded by other single-item trials, to minimize the risk of any potential memory task errors influencing the proceeding critical items and so that these were not the only single-item ones. Participants made very few errors on the memory task overall and were almost all at ceiling-level performance.

On nearly 50% of the trials (73% of the multiple-item trials), the words were semantically related. The critical items were presented in the order: *atleet*, *schaduw*, *stempel*, and *saldo*,

[4] These two sounds are realized differently, sometimes inversed, in many regional accents of Dutch (Collins & Mees, 2003).

[5] The speaker did this inconsistently and may have transferred it from German.

[6] Note that, according to Collins and Mees (2003), the Dutch [s] is produced with the blade of the tongue and in many regions of The Netherlands (e.g., the Randstad) is produced more lax and/or retracted, often sounding graver, like [ʃ], especially in clusters, word-finally, and after 'r' (Ditewig, Pinget, & Heeren, 2019).

which repeated five times for a total of five utterances of each word. There were four filler trials between critical items. In order to examine convergence over time, only the first and fifth utterances of each word were analyzed. Tokens two-four were not annotated nor analyzed due to limitations in length of the AXB rating study in favor of being able to get each token rated by all raters.

Before beginning the task, participants did five practice trials. The repetition task took about 10 minutes overall. Immediately after the memory task, the participants were probed for any suspicions of the cover story or awareness of convergence. Specifically, they were asked what they thought the study was about, whether they noticed anything about the model speaker, and whether they noticed themselves imitating the speaker, either intentionally or not. About two-third of participants believed the task was about memory. The most commonly cited study purpose concerned the role of the semantic relationship between some of the words in memory, but other memory explanations included having to recall the words from the baseline word reading task or memory task later, and speculations about the effect on memory of repeating words in reverse order, sequence length, the number of repetitions, the order in which a word appeared in a sequence, the phonological similarity between words (e.g., *neus*, *fornuis*), the frequency of words, or the number of syllables. Crucially, only three participants (3.2%) indicated that they thought the task might have something to do with convergence and seven more mentioned pronunciation (7.5%). Excluding these participants did not change the pattern of results so analyses reported include them. Interestingly, when inquired, an even larger number of participants (nearly 40%) reported noticing themselves imitate the model's speech in one way or another.

Following these questions, participants were debriefed and fully informed about the purpose of the study and what would be done with the recordings of their utterances, as per regulations regarding studies involving deception. They were asked to provide written consent again if they wished to continue.

Participants were then asked some questions about the model speaker's accent: whether they thought she was a native speaker of Dutch (and how certain they were of their response); if so, if they had to guess, what region they would say the model was from and, if not, what they thought her mother tongue was (and how certain they were of their response). They were then asked to rate the model speaker's accent in terms of how strong, comprehensible, and familiar they found it. The accentedness and comprehensibility scales were adapted from Munro and Derwing's work (e.g., Munro & Derwing, 1999), and the familiarity scale from Witteman, Weber, and McQueen (2013). Participants' responses to these questions are reported in 2.3.1.

On the second day of the experiment, after all of the other tasks were completed, participants filled out a questionnaire with demographic and language background questions. This questionnaire included questions about their regional accent in Dutch and their familiarity with Croatian and Croatian-accented Dutch.

2.4.3. AXB perceptual similarity task

On each trial of the AXB perceptual similarity task, raters heard three utterances of the same word: the model speaker's utterance (X) the participant heard during the convergence task, the participant's repeated utterance and the participant's baseline production of the same word before exposure to the model (A and B). The raters' task was to decide which of the participant's utterances sounded more like the model speaker's. If phonetic convergence is perceived, the raters should select the participants' repeated utterances more often than the participants' baseline utterances.

The speaker's baseline items were the utterances obtained from the word reading task before the memory task. In all cases except one where the recording was not useable, the participant's first reading of the word was used for the baseline utterances. The participants' first and fifth tokens from the convergence task were used for the repetitions. Word boundaries of the baseline and repeated items were manually annotated for cropping in Praat (Boersma & Weenink, 2018) and extracted from the audio files with a script. Model and participant recordings were amplitude-normalized to have the same overall RMS (root-mean-square) value (the average for the audios), all converted to 48 kHz-sampling frequency, and concatenated using Praat scripts. Eight experimental lists were created, Latin square-counterbalancing for the order of the four critical items and for the position (A or B) of the repeated item so that it appeared in each position half of the time per list. Items were randomized within word per participant, with six "anchors" (participant data otherwise excluded) at the beginning of each word for raters to get an idea of the range in performance (cf. Kim et al., 2011). Trials were divided into three blocks per word (after each 56 trials) to allow raters a break. The rating task was administered via LimeSurvey.

Raters performed the task in the lab and were randomly assigned to one of the eight lists, with each rater hearing all items. The instructions were to indicate which recording sounded more similar in pronunciation to the middle one (X): the first (A) or the last (B) one. The experiment was self-administered and the raters were allowed to replay the audios as many times as necessary. They provided their responses by selecting the option "A" or "B". The task began with three practice trials with one of the practice items used in the memory task. In total, each rater performed 675 AXB trials (4 words $\times$ 2 tokens $\times$ 81 participants + 6 fillers $\times$ 4 words + 3 practice items). The whole rating experimental session lasted on average about 90 minutes (range: 60–135).

2.5. Data analysis

The data described here are available from the Donders Repository (data.donders.ru.nl) and can be found via the collection identifier: di.dcc.DSC_2017.00132_627, or persistent identifier: https://doi.org/10.34973/y7x7-ct33. The data can be accessed and downloaded upon registration and acceptance of the data use agreement.

2.5.1. Pre-processing of acoustic data

Sound and word boundaries of critical, reference, and vowel space items were manually annotated in Praat. All acoustic measures were obtained from Praat. Annotations were carried out with auditory and visual inspection of the waveform and spectrogram. All boundaries were placed at the nearest zero-crossing using a script. The data was pre-processed and ana-

lyzed in R (version 4.0.2; R Core Team. (2020), 2020). The preprocessing steps described here were applied to both the participants' and the model speaker's utterances.

All measures of sound duration (e.g., vowel, aspiration, fricative) were divided by word duration, becoming relative duration measures (i.e., proportion of word duration). This was done in order to account for differences in speech rate, which could have especially affected the differences in the two tasks: word reading and repetition. Table 1 provides a summary of all of the acoustic dimensions measured per word.

2.5.1.1. Sound-specific measures. [t]. In final position, the durations of the stop closure and vowel, which have been found to be negatively related, become more relevant to distinguish stops in Dutch (Kuijpers, 1993; Slis & Cohen, 1969), which has final devoicing. Thus, we measured closure duration from the end of the preceding vowel's periodicity to the start of the stop's release burst. Aspiration duration was also analyzed however, measured from the start of the release burst until the end of the burst or aspiration, if present. The baseline aspiration data of two participants had to be excluded due to coarticulation with the following word. Furthermore, stop aspiration and closure measures for three participants who produced the final consonant sound in *atleet* as the affricate [ts] were excluded from the analysis.

[o:]. Since [o:] was a diphthong and word-final, sound offset was set at word offset. Sound onset was set at the onset of periodicity. Duration and formant (F1 and F2) values were automatically extracted with a script. In order to compare between different speakers, F1 and F2 were normalized for anatomical differences in the speakers' vowel space by applying the Lobanov transformation (Adank et al., 2004; van der Harst, 2011) using the phonR package (version 1.0–7; (McCloy & McCloy, 2016). The formant values for one participant were excluded due to creaky voice which impeded measurement.

Two indices of diphthongization were measured: vowel endpoint and movement. Vowel endpoint was measured as F1 and F2 at 75% of the vowel's duration, while movement was the difference between the beginning (25%) and end (75%) of the vowel. For that, formant values were also extracted at 25% of the vowel's duration and the Euclidean distance between the two points was calculated as follows: $\sqrt{((F1_{75\%} - F1_{25\%})^2 + (F2_{75\%} - F2_{25\%})^2)}$.

[a:]. Sound boundaries were set at the onset and offset of periodicity. Vowel duration and formant values of the vowel's temporal midpoint were automatically extracted using a script. F1 and F2 were normalized with the Lobanov transformation, as for [o:].

[s]. In most languages, the fricatives [s] and [ʃ] vary in terms of spectral center of gravity (CoG), with [s] having a longer articulatory tract and thus higher CoG than [ʃ] (Gordon, Barthmaier, & Sands, 2002). Rietveld and van Heuven (2009) have also shown this for Dutch. The alveolar [s] also tends to have a longer duration than postalveolar [ʃ], although duration has been found to be less important in distinguishing voiceless fricatives (Ditewig, Pinget, & Heeren, 2019; Gordon et al., 2002). Considering this, we analyzed CoG, as well as duration of the fricative. The fricative was annotated with pre-emphasis set to 0 and dynamic range to 50. Boundaries were set at the start and end of frication noise. Then, the CoG of the center 50% of the fricative's spectral slice and fricative duration was automatically extracted using scripts.

Table 1
Summary of acoustic measures used for each item.

Word	Critical sound	Acoustic measures
atleet	[t]	stop closure duration (relative) stop aspiration duration (relative) speech rate (word duration) f0 (median)
saldo	[o:]	vowel movement vowel endpoint vowel duration (relative) speech rate (word duration) f0 (median)
schaduw	[a:]	vowel midpoint vowel duration (relative) speech rate (word duration) f0 (median)
stempel	[s]	fricative CoG fricative duration (relative) speech rate (word duration) f0 (median)

2.5.1.2. General measures. In addition to the above sound-specific measures specially selected to measure convergence to the model speaker's realizations, speech rate and f0, dimensions often analyzed in studies on convergence to natives, were also measured for all critical words.

Speech rate. Because the items were single words, speech rate was measured as raw word duration.

F0. Fundamental frequency (f0) values were extracted with a script that used Praat's default range of 75–600 Hz and a 10 ms-step. A 20 ms-buffer was added to word boundaries for the pitch time window analysis. Unvoiced segments, whose f0 values returned undefined values, were excluded. F0 values below 110 Hz, which are thought to reflect creaky voice, were discarded and values that were more than double or less than half than the previous value were checked for errors in Praat's formant contour tracking (cf. Marcoux & Ernestus, 2019). Each participant's median f0 was calculated per utterance.

2.5.2. Difference-in-distance (DID) scores

Convergence was assessed acoustically with difference-in-distance (DID) scores (see Pardo et al., 2013). First, distance measures for all items were obtained by calculating the absolute difference between the model speaker's values and those of each participant's value. This was done for both baseline (a) and repeated (b) items and for every acoustic dimension. For vowel spectra, Euclidean distances were calculated with the following formula: $\sqrt{((F1_{participant} - F1_{model})^2 + (F2_{participant} - F2_{model})^2)}$ (Bradlow, Torretta, & Pisoni, 1996; Pardo et al., 2017). This way, F1 and F2 were combined into one measure following claims that treating them as two-dimensional points is more robust and more valid than analyzing them separately (e.g., Pardo et al., 2017; Vallabha & Tuller, 2004).

Next, DID scores were calculated by subtracting each repeated distance from the participant's baseline distance. For example, let's take the [a:] in *schaduw*, which the model speaker produced with a relative duration of 0.28 of the word. If a participant produced the sound with a relative duration of

0.20 during baseline and 0.25 during the memory task, their DID score for that token would be |0.20–0.28|-|0.25–0.28| = 0 .05. In this way, positive values indicate convergence, whereas negative values indicate divergence and differences of 0 indicate no change from baseline to repetition. The amplitude of the DID score reflects the amplitude of the change from baseline (Lewandowski & Nygaard, 2018).

Outliers greater than 2.5 SD from the mean DID were removed from the data of each acoustic dimension. This resulted in a total exclusion of 51 data points overall in the acoustic analyses (2.0% of the data; 2.4% for *atleet*, 2.1% for *saldo*, 1.5% for *schaduw*, and 2.0% for *stempel*) and 43 data points in the analyses of the relationship between the perceptual and acoustic measures (7.3% of the data; 7.6% for *atleet*, 8.6% for *saldo*, 6.0% for *schaduw*, and 6.8% for *stempel*).

2.5.3. Statistical analyses

Following Pardo et al. (2013), phonetic convergence was assessed using linear mixed-effects models (MEMs) with the packages lme4 (version 1.1–23; Bates et al., 2015) and lmerTest (version 3.1–2; Kuznetsova, Brockhoff, & Christensen, 2017) for p-values. Perceived convergence was analyzed using generalized linear mixed-effects models of the rating data. Models employed maximal random effect structures (participant, rater, token), including random intercepts and slopes where appropriate (Barr, Levy, Scheepers, & Tily, 2013). The inclusion of each fixed effect was justified by ascertaining that it improved fit compared to a model without it using Chi-square tests (Baayen, Davidson, & Bates, 2008; Jaeger, 2008).

Responses from the AXB perceptual similarity task were recoded as 0 (baseline more similar to model) and 1 (repetition more similar to model). All categorical predictors were contrast-coded (−0.5, 0.5) in the following orders: A vs. B (position) and 1 vs. 5 (token). Simple effects coding was used for word in the order *atleet*, *saldo*, *schaduw*, *stempel*. A parameter for the item's position in the trial (A vs. B) was included to account for raters' bias in choosing one over the other in the AXB task.

In order to be able to compare the contribution of the different numeric predictors, all DID scores and model ratings were scaled and centered by conversion into z scores (Pardo et al., 2017). Multicollinearity between predictors was evaluated using Cramér's V for categorical variables (i.e., position and token; vcd, Zeileis, Meyer, & Hornik, 2007), intra-class correlations for categorical and continuous variables (ICC, Wolak, Fairbairn, & Paulsen, 2012), and kappa κ for multiple continuous variables (collin.fnc in languageR, Baayen & Shafaei-Bajestan, 2019). In every case, association between variables was weak, suggesting lack of collinearity.

3. Results

3.1. Perceived convergence

With responses to the AXB task recoded as 0 (baseline selected) and 1 (repetition selected), the average gives the proportion of trials on which the repeated response was chosen. Fig. 1 shows the results of the AXB rating study by word and token. If raters were to respond randomly, we would expect an average response rate of 0.50 (reference line in Fig. 1). Studies finding perceived convergence tend to observe subtle effects, with average values around 0.56 (Pardo et al., 2018). Here the overall AXB phonetic convergence averaged 0.58, which was significantly greater than chance as indicated by the significance of a null model with only random intercepts (participant and rater; $\beta = 0.362$, $z = 4.064$, $p < 0.001$).

Adding a fixed effect for word improved the model's fit ($\chi^2(3) = 49.880$, $p < 0.001$), suggesting that raters perceived different levels of convergence depending on the word being rated ($\beta_{saldo} = 0.119$, $\beta_{schaduw} = -0.337$, $\beta_{stempel} = 0.584$). Changing the reference level (Šidák-corrected $\alpha = 0.017$ for multiple comparisons) revealed that raters perceived less convergence to *schaduw* (0.49) than to *atleet* (0.56), *saldo* (0.59), and *stempel* (0.68), and more convergence to *stempel* than to *atleet* and *saldo*.

Adding a fixed effect for token (1 vs. 5) significantly improved the model's fit further ($\beta = -0.154$, $z = -2.593$, $p = 0.010$; $\chi^2(1) = 6.463$, $p = 0.011$). Overall, raters perceived more convergence on the first token (0.59) than on the fifth and last one (0.56). There was no significant overall interaction between word and token ($\chi^2(3) = 6.243$, $p = 0.100$), although numerically the token effect seemed to only be present for *atleet* and *saldo*, as can be seen in Fig. 1.

Next, the participants' ratings of the model speaker's accent were included in the model. Note that only participants who perceived the model speaker as non-native are included so that the minimum accentedness rating was 2 rather than 1 (no foreign accent/nativelike). Only the addition of perceived accentedness improved the model's fit ($\beta = -0.134$, $z = -2.821$, $p = 0.005$; $\chi^2(1) = 7.564$, $p = 0.006$), with raters perceiving less convergence for speakers who rated the model speaker's speech as more accented. Perceived comprehensibility ($\chi^2(1) = 0.183$, $p = 0.669$) and familiarity with the accent ($\chi^2(1) = 0.011$, $p = 0.917$) were not found to have a significant effect on degree of convergence. Additionally, no interaction was observed between perceived accentedness and word ($\chi^2(3) = 0.485$, $p = 0.922$) nor token ($\chi^2(1) = 1.730$, $p = 0.188$).

These results suggest that raters perceived convergence overall, although this varied by word, with the greatest degree of perceived convergence for *stempel* and the least for *schaduw*. Moreover, the results of the perceptual analyses suggest that convergence to the model speaker decreased over the task, with greater perceived convergence for the first repetition compared to the fifth. Finally, the ratings suggest that speakers tended to converge less the stronger they perceived the model's accent to be, while perceived comprehensibility and familiarity did not seem to play a role.

3.2. Acoustic analysis of convergence

Table 2 presents an overview of the model speaker and participants' values for each one-dimensional acoustic measure with the dimensions in which the model speaker's values were outside the participants' baseline range presented in bold. Fig. 2 displays participants' and the model speaker's two-dimensional spectral values for *saldo* and *schaduw*.

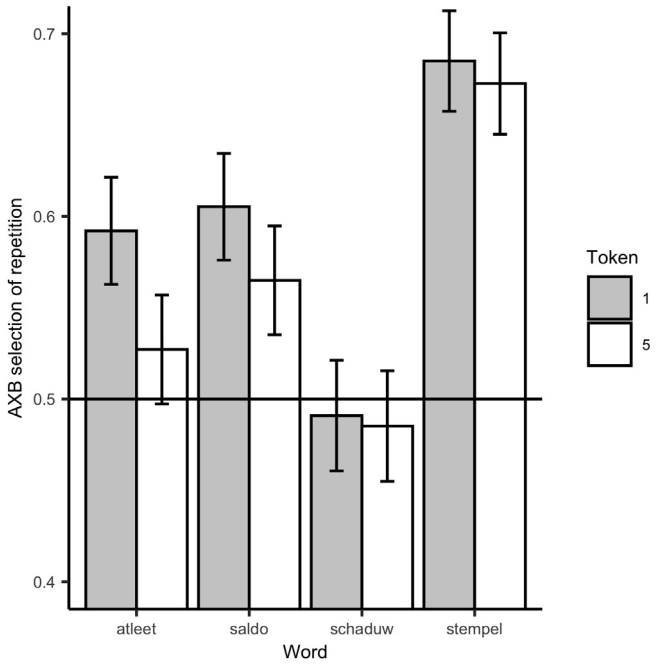

Fig. 1. Average perceived convergence (AXB) per word and token. Error bars represent 95% confidence intervals.

Convergence on acoustic dimensions was assessed via linear mixed-effects models (MEMs) carried out individually per dimension. As was done for the rating data, convergence on the acoustic dimensions was determined by Chi-square tests of the model against the null model with only random effects (for speaker). A summary of the results of the acoustic analyses can be found in the last column of Table 2.

Atleet. As expected because of its position in the word, the relative duration of the model speaker's aspiration (0.20) was not outside of the native Dutch participants' baseline range (M = 0.24, SD = 0.05). The model speaker's values were farther from the native speakers' average values for the remaining acoustic dimensions measured: stop closure duration (model: 0.14; participants: M = 0.08, SD = 0.02), speech rate (model: 808 ms; participants: M = 644 ms, SD = 65), and f0 (model: 233 Hz; participants: M = 202 Hz, SD = 21). Accordingly, the results of the null MEMs for *atleet* revealed significant overall convergence for stop closure duration (β = 0.005, t = 2.492, p = 0.015), speech rate (β = 17.263, t = 2.764, p = 0.007), and f0 (β = 2.941, t = 2.218, p = 0.030). In addition, there was an effect of token for speech rate (β = −19.612, t = −2.864, p = 0.006; $\chi^2(1)$ = 7.808, p = 0.005), with participants converging more on the first (M = 671 ms, SD = 71; DID = 26.60 ms) as opposed to the fifth token (M = 652 ms, SD = 73; DID = 7.69 ms). Furthermore, there was an interaction between token and perceived comprehensibility for f0 (β = −3.003, t = −2.416, p = 0.019; $\chi^2(3)$ = 8.537, p = 0.036), with greater perceived comprehensibility leading to greater convergence on the first token but not the last. In addition, convergence to f0 was modulated by perceived familiarity with the model speaker's accent (β = 3.271, t = 2.494, p = 0.015; $\chi^2(1)$ = 6.236, p = 0.013), with participants converging more the more familiar they perceived the accent to be. Overall there was no significant convergence to stop aspiration (β = 0.0004, t = 0.014, p = 0.913). However, there was a significant effect of perceived accentedness for aspiration DID (β = −0.010, t = −2.429, p = 0.018; $\chi^2(1)$ = 5.813, p = 0.016), with slightly less convergence and even divergence the stronger participants rated the model's accent.

Saldo. For this word, the model speaker's values were fairly extreme for all dimensions, but only outside of the participants' baseline range for the measures corresponding to the realization of the [oː]: vowel movement (model: 0.29; participants: M = 1.05, SD = 0.32; for vowel endpoint, see the spectral plot in Fig. 2) and vowel duration (model: 0.28; participants: M = 0.41, SD = 0.04). The model speaker's speech rate (model: 751 ms; participants: M = 698 ms, SD = 74) and f0 (model: 235 Hz; participants: M = 205 Hz, SD = 21) were closer to the average values, although her f0 was still relatively high. The within-range speech rate was likely due to the model's reduced duration of the final [oː] vowel as for the other items her speech was slower than that of the participants. Interestingly, the results of the null MEMs only found that DID significantly varied from 0 for vowel duration and speech rate. Overall the participants converged to the model speaker's extremely short vowel duration (β = 0.026, t = 6.736, p < 0.001), and diverged from her speech rate (β = −29.681, t = −4.280, p < 0.001), producing the word faster than at baseline. Again, this could be due to the reduced vowel duration participants converged to. Participants did not seem to converge to the model's decreased vowel movement (β = −0.017, t = −0.452, p = 0.652) nor spectrally to the vowel's endpoint (β = −0.001, t = −0.021, p = 0.983). Finally, no convergence was observed for the model speaker's slightly elevated f0 (β = 0.740, t = 0.355, p = 0.723), nor was token nor the participants' ratings of the model speaker's accent found to play a role.

Schaduw. As can be seen in Fig. 2 and Table 2, the model's realization of [aː] fell outside of the native participants' mean range spectrally, approaching [ɑ], although native-like in terms of relative duration (model: 0.28; participants: M = 0.26, SD = 0.03). The model speaker's speech rate was much slower than that of the participants (model: 812 ms; participants: M = 648 ms, SD = 77), as for almost all of the other items. The model speaker's median f0 value for this word was very close to the participants' mean (model: 189 Hz; participants: M = 186 Hz, SD = 21), although relatively low for the model speaker (188 Hz vs. 233–243 Hz for the other critical items) whose f0 tended to be higher than the participants'. This is likely due to the fact that the utterance of this word was taken from the end of a list, giving it a salient, atypical intonation. The results yielded by the null MEMs for *schaduw* revealed that participants significantly converged overall to speech rate (β = 41.510, t = 6.115, p < 0.001) and diverged in f0 (β = −4.248, t = −2.632, p = 0.010), increasing their f0 even past that of the model. Furthermore, the addition of perceived comprehensibility increased model fit for both of these dimensions (speech rate: β = −15.799, t = −2.272, p = 0.026; $\chi^2(1)$ = 5.126, p = 0.024; f0: β = −4.371, t = −2.688, p = 0.009; $\chi^2(1)$ = 7.077, p = 0.008). For speech rate, this translated to more convergence with lower perceived comprehensibility and for f0: greater divergence the greater the rated comprehensibility. No overall convergence was observed to

Table 2
Summary of model speaker and participant values (baseline and repetition, across both attempts) for each one-dimensional acoustic measure and of results of acoustic analyses of convergence.

				Participants				
				Baseline		Repetition		Accommodation pattern
Word	Critical sound	Acoustic measure	Model speaker	M (SD)	Range	M (SD)	Range	
atleet	[t]	stop closure duration (relative)	0.14	0.08 (0.02)	0.02–0.18	0.08 (0.03)	0–0.16	convergence
		stop aspiration duration (relative)	0.20	0.24 (0.05)	0.10–0.36	0.24 (0.04)	0.11–0.35	maintenance
		speech rate (word duration, ms)	**807.66**	**644.20 (64.60)**	**469.56–802.33**	661.21 (72.12)	482.74–802.14	convergence (token 1 > 5)
		f0 (median, Hz)	232.87	202.21 (20.76)	152.36–267.43	205.91 (21.68)	155.15–261.96	convergence
saldo	[oː]	**vowel movement**	**0.29**	**1.05 (0.32)**	**0.36–2.03**	1.05 (0.33)	0.08–1.72	maintenance
		vowel endpoint	*See* Fig. 2					maintenance
		vowel duration (relative)	**0.28**	**0.41 (0.04)**	**0.30–0.49**	0.38 (0.04)	0.25–0.49	convergence
		speech rate (word duration, ms)	751.14	698.23 (73.57)	560.13–863.41	654.56 (76.85)	513.67–906.31	convergence
		f0 (median, Hz)	234.62	204.77 (21.04)	153.21–272.43	210.66 (27.30)	152.84–279.10	maintenance
schaduw	[aː]	vowel midpoint	*See* Fig. 2					maintenance
		vowel duration (relative)	0.28	0.26 (0.03)	0.19–0.32	0.26 (0.03)	0.19–0.34	maintenance (token 5 > 1)*
		speech rate (word duration, ms)	811.65	648.22 (76.66)	512.16–922.21	690.36 (78.50)	525.02–901.23	divergence
		f0 (median, Hz)	188.84	185.93 (20.85)	143.38–255.61	201.07 (23.23)	141.61–262.01	divergence
stempel	[s]	**fricative CoG (Hz)**	**3831.05**	**5700.00 (574.70)**	**4748.48–7020.33**	5541.61 (633.49)	4384.60–7421.96	convergence
		fricative duration (relative)	0.20	0.18 (0.03)	0.12–0.26	0.17 (0.03)	0.10–0.24	divergence (token 5 > 1)
		speech rate (word duration, ms)	799.81	620.57 (87.64)	423.99–855.39	658.23 (81.58)	498.30–916.00	convergence
		f0 (median, Hz)	243.48	190.19 (22.46)	141.83–252.57	216.28 (31.26)	146.26–301.47	convergence

Note: interactions with participants' ratings of the model speaker's accent are not included here for simplicity, but see details in Section 3.2.
*Convergence on token 5, divergence on token 1.

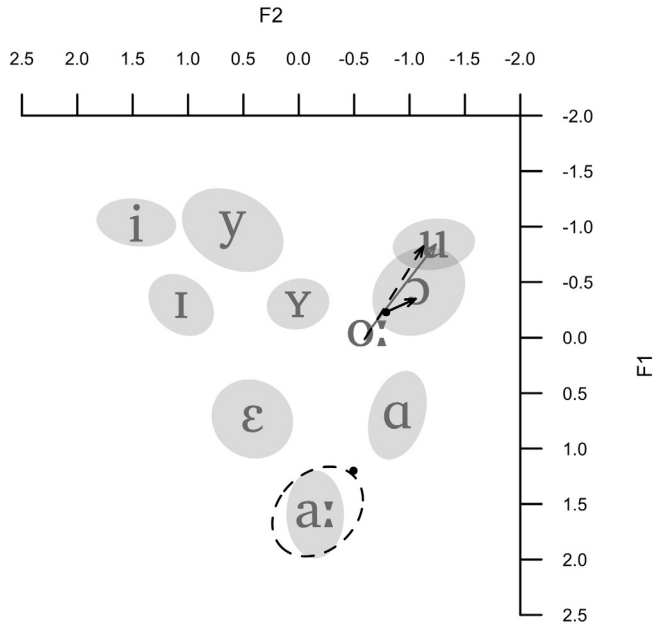

Fig. 2. Spectral plot of participants' and model speaker's utterances of *schaduw* and *saldo*. Participants' baseline values are shown in light gray, critical item repetitions are represented with dark gray dashed lines, and the model speaker's realizations are indicated with the black arrow and filled circles. Values from the critical words *schaduw* and *saldo* are plotted for [aː] and [oː] and reference words *schatkist* and *haardos* were used for [ɑ] and [ɔ]. Arrows represent vowel movement (25–75%). Ellipses represent 1 SD (68% CI).

the two dimensions of the vowel measured (midpoint: $\beta = -0.011$, $t = -0.536$, $p = 0.593$; vowel duration: $\beta = 0.001$, $t = 0.630$, $p = 0.531$). However, a complex picture emerged for vowel duration. Adding token improved model fit ($\beta = 0.008$, $t = 3.809$, $p < 0.001$; $\chi^2(1) = 13.941$, $p < 0.001$), with participants diverging in duration on the first token ($M = 0.26$, $SD = 0.03$; DID = -0.003) and converging on the last ($M = 0.27$, $SD = 0.02$; DID = 0.006). Moreover, perceived comprehensibility was found to interact with token ($\beta = -0.005$, $t = -2.370$, $p = 0.020$; $\chi^2(2) = 6.795$, $p = 0.033$), with participants who rated the model speaker as more comprehensible converging less and even diverging only on the fifth token. Perceived comprehensibility was further found to interact with how strong the participants rated the model speaker's accent ($\beta = 0.007$, $t = 2.122$, $p = 0.037$; $\chi^2(2) = 7.199$, $p = 0.027$), with participants converging more to vowel duration with greater perceived accentedness when they perceived the model speaker as highly comprehensible and conversely diverging more with greater accentedness when they perceived the model as less comprehensible. Finally, familiarity was also found to improve model fit ($\beta = -0.005$, $t = -2.197$, $p = 0.031$; $\chi^2(1) = 5.001$, $p = 0.025$), as participants appeared to converge less and even diverge more the more familiar they were with the model speaker's perceived accent.

Stempel. The model's values were extreme for almost all of the dimensions, namely: CoG (model: 3831 Hz; participants: $M = 5700$ Hz, $SD = 575$), speech rate (model: 800 ms; participants: $M = 620$ ms, $SD = 88$), and f0 (model: 243 Hz; participants: $M = 190$ Hz, $SD = 22$). For fricative duration, the model's value of 0.20 ms was near participants's mean ($M = 0.18$, $SD = 0.03$). However, the results of the MEMs for *stempel* suggested that participants adapted their production from baseline for all of the acoustic dimensions measured. Participants

converged to the model speaker's lower CoG (β = 158.39, t = 3.07, p = 0.003), slower speech rate (β = 33.425, t = 4.353, p < 0.001) and elevated f0 (β = 18.729, t = 8.659, p < 0.001). Conversely, the participants diverged from the model speaker on the dimension that was within the native range, fricative duration (β = −0.008, t = −3.011, p = 0.004), producing it shorter. Moreover, addition of a fixed effect for token improved model fit for fricative duration (β = −0.011, t = −4.030, p < 0.001; $\chi^2(1)$ = 14.849, p < 0.001), with participants diverging more on the fifth token (M = 0.16, SD = 0.03; DID = −0.014) than the first (M = 0.18, SD = 0.03; DID: −0.003). Nor did convergence vary by token for any of the other dimensions, nor were the participants' perceptions of the model speaker's accent related to degree of convergence for any measures of this item.

3.3. Relationship between acoustic and perceptual measures of convergence

In order to examine the relationship between acoustic and perceptual measures of convergence, separate MEMs were run for each word with the AXB rating data as the dependent measure and the acoustic measures as the predictors. Participant and rater random effects were included, as well as a fixed effect to control for effect of position. Here, DID scores were transformed into z-scores to be able to compare the contribution of the different acoustic dimensions to AXB ratings. A summary overview of the results can be found in Table 3.

Atleet. The inclusion of DID for stop closure duration (β = −0.189, z = −2.489, p = 0.013), speech rate (β = 0.313, z = 4.108, p < 0.001), and f0 (β = 0.237, z = 2.530, p = 0.011) significantly improved model fit compared to the null model ($\chi^2(4)$ = 32.000, p < 0.001). Inspection of the beta estimates suggests greater perceived convergence to greater convergence for f0 and speech rate, while greater perceived convergence was related to greater divergence for closure duration. Stop aspiration duration DID, which did not display overall convergence in the acoustic models, was also not found to be related to perceived convergence ($\chi^2(1)$ = 1.069, p = 0.301).

Saldo. MEMs on the AXB ratings for this word revealed that adding the DID of vowel duration (β = 0.251, z = 3.496, p < 0.001), speech rate (β = 0.178, z = 2.488, p = 0.013), and f0 (β = 0.354, z = 4.325, p < 0.001) all significantly improved model fit ($\chi^2(4)$ = 83.868, p < 0.001) and were positively related to perceived convergence. Vowel movement ($\chi^2(1)$ = 0.018, p = 0.894) and vowel endpoint ($\chi^2(1)$ = 2.766, p = 0.096) DID, which did not reveal overall convergence, were also not found to be related to perceived convergence.

Schaduw. Adding vowel duration (β = 0.160, z = 2.628, p = 0.009), speech rate (β = 0.335, z = 5.218, p < 0.001), and f0 (β = 0.327, z = 5.052, p < 0.001) to the model for ratings of *schaduw* all significantly improved model fit ($\chi^2(4)$ = 88.109, p < 0.001). Furthermore, the beta estimates were all positive, indicating greater perceived convergence to greater convergence along these acoustic dimensions. Only vowel quality (F1 and F2; $\chi^2(1)$ = 0.141, p = 0.708), which participants did not converge to, was not found to be related to raters' judgments of convergence.

Table 3
Summary of significant results of analyses of relationship between perceptual and acoustic analyses of convergence.

Word	Critical sound	Acoustic measure	Relationship with AXB ratings
atleet	[t]	stop closure duration (relative)	negative
		stop aspiration duration (relative)	–
		speech rate (word duration)	positive
		f0 (median)	positive
saldo	[oː]	vowel movement	–
		vowel endpoint	–
		vowel duration (relative)	positive
		speech rate (word duration)	positive
		f0 (median)	positive
schaduw	[aː]	vowel midpoint	–
		vowel duration (relative)	positive
		speech rate (word duration)	positive
		f0 (median)	positive
stempel	[s]	fricative CoG	–
		fricative duration (relative)	–
		speech rate (word duration)	positive
		f0 (median)	positive

Stempel. Despite participants adapting to all acoustic dimensions measured for *stempel*, only convergence to speech rate (β = 0.320, z = 4.112, p < 0.001) and f0 (β = 0.315, z = 4.037, p < 0.001) were found to be related to perceived convergence ($\chi^2(3)$ = 47.727, p < 0.001), with greater acoustic convergence revealing greater perceived convergence, as well. Convergence to CoG ($\chi^2(1)$ = 1.284, p = 0.257) and divergence to fricative duration ($\chi^2(1)$ = 1.572, p = 0.210) were not found to be related to the degree of convergence perceived by the raters.

Overall, the results of these analyses suggest that raters picked up on the acoustic dimensions participants converged to. Exceptions are CoG and fricative duration for *stempel*, which were adapted to but were not found to add to raters' judgments, and f0 for *saldo*, which raters seemed to make use of despite participants not converging to this dimension overall. Across words, f0 and speech rate were always found to be related to degree of convergence perceived by the raters, along with all other durational measures except fricative duration (*stempel*) and stop aspiration (*atleet*), the latter of which was not converged to.

4. Discussion

The present study aimed to examine whether native speakers can display phonetic convergence to non-native speech in a non-interactive setting, in which socio-motivational influences resulting from in-person interaction are minimized. The results reveal that, overall, native participants converged to a non-native speaker with an unfamiliar accent in a repetition task disguised as a memory task. In addition, convergence was apparent in both perceptual and acoustic measures and there was a relationship between some of the acoustic dimensions measured and perceived convergence as assessed by an AXB perceptual similarity task. Moreover, as is commonly observed in the accommodation literature (e.g., Levitan & Hirschberg, 2011; Pardo et al., 2017, 2018), participants converged in varying degrees to different items, tokens, and

acoustic dimensions. In addition, the rating results suggest that the stronger participants rated the non-native speaker's accent to be, the less likely they were to converge. In contrast, perceived comprehensibility and familiarity with the accent only played a role in convergence on some acoustic dimensions.

4.1. Perceived convergence

The main goal of this study was to evaluate phonetic convergence to non-native speech. To that end, perceptual ratings are usually considered a more holistic measure, integrating convergence across different acoustic dimensions. The results of the AXB rating task reveal that, at a rate greater than that expected by chance, participants' memory task utterances were chosen over their baseline productions as being more similar to the critical items produced by the model speaker, suggesting that overall the participants converged to the non-native model speaker. The average level of perceived convergence observed here (0.58) was comparable to those reported in studies on convergence in general (0.56; Pardo, Urmanche, Wilman, et al., 2018) and, crucially, to the two other studies of convergence to non-native speech in non-interactive settings (i.e., 0.55 for Kim, 2012; 0.57 for Lewandowski & Nygaard, 2018).

The degree of perceived convergence varied across items, with raters on average detecting convergence to all words except *schaduw,* likely due to its atypically low intonation, and the most convergence to *stempel*. The latter might make sense in light of the fact that the word *stempel* was the only one to reveal convergence across all acoustic dimensions measured. However, only speech rate and f0 were found to be related to raters' perception of convergence. An alternative explanation might relate to the model's utterance rising intonation, which was partly captured by the elevated measure of f0. If participants imitated the token's intonation, that would likely have been especially salient in the perceptual similarity rating task. The word was salient to the participants of the memory task as it was the word most remarked upon during debriefing. It has been proposed that listeners converge more to perceptually salient features (Levitan, 2020; Walker & Campbell-Kibler, 2015; but see Babel, 2010, who suggests the opposite). Thus, the rising intonation could explain the increased perceived convergence for this item.

The results of the AXB perceptual analyses also indicated greater convergence to the first compared to the fifth token overall, although closer inspection suggests this might only have been the case for *atleet* and *saldo*. Our findings are consistent with previous reports of convergence occurring early on, but the inconsistent token effect seems to contradict the idea that convergence increases over time, as exposure to the target speech increases (e.g., Pardo, 2006). However, it is worth noting that in our study, order is confounded with repetition of the items, with each item repeated five times. In this respect, our findings still contradict Goldinger's (1998) observation that degree of perceived convergence increased with number of exposures, as well as Lewandowski and Nygaard's (2018) speculation that increased exposure and experience with an accent may help to lift the inhibiting influence that non-native speech may have on convergence. However, order effects have proven to be fairly inconsistent in the literature (Babel, 2012; Pardo, Urmanche, Wilman, et al., 2018), while repetition effects are often avoided. Considering that our token effect does not seem to be very stable across items, and given the limited set of items used here, further research is needed to understand how and why order and repetition effects may manifest under these circumstances.

In the present study, a non-interactive task was employed in order to evaluate convergence in a context minimizing the potential influence of factors resulting from in-person interaction, such as attitudes to the other speaker resulting from their physical appearance or from the interaction itself, as well as use of communicative strategies. At the same time, the general assumption that native speakers do not seek to sound like non-native speakers in their mother tongue, and non-native accents often carry a social stigma (Gluszek & Dovidio, 2010), led us to assume that any convergence observed would not be the result of the participants wanting to sound like the model speaker. To test this, we included the participants' ratings of the model speaker's accent in the analyses.

The fact that we observed convergence to what the participants recognized as non-native speech suggests the model's status as an L2 speaker did not completely block the tendency to converge; native speakers can still converge to another speaker despite their non-native accent and the lack of any socio-communicative motivation to do so. However, the AXB rating results suggest that tendency to converge was modulated by perceived accentedness, with participants converging less the stronger they perceived the non-native speaker's accent to be. This lends support to the idea that participants' tendency to converge is still mitigated by social effects such as perceived accentedness, even in a non-interactive task (cf. Babel, 2010).

In contrast, here we did not find perceived comprehensibility to play an overall role in the degree to which participants were perceived to converge. Previous studies have suggested that native speakers may phonetically converge to non-native interlocutors in order to facilitate comprehension (e.g., Costa et al., 2008; Kim, 2012). However, as discussed before, this effect may be specific to interactive settings where alignment can serve to improve communication. Despite this, Lewandowski and Nygaard (2018), in their auditory naming study with native and non-native speakers, found greater perceived convergence to the non-native speakers, which they attributed to the greater perceived accentedness and lower intelligibility of the non-native speakers. Since the ratings of the model speakers in Lewandowski and Nygaard's (2018) study were from a different set of raters, they could not be included in the analyses and their effects directly evaluated. Although untested, the authors also suggested that experience with the non-native speech during a perceptual exposure task before performing auditory naming may have reduced any potential negative attitude towards the non-native speech which could have blocked convergence. However, in the present study, in which we included participants' own ratings of the model speaker in the analyses, we found that perceived accentedness decreased the tendency to converge to our non-native speaker, while perceived comprehensibility and familiarity with the accent did not have an effect overall. These results are also not in line with the idea that increased attentional

demands and processing load in perceiving non-native speech may block the automatic tendency to converge (Costa et al., 2008; Kim et al., 2011). If this were the case, we would expect raters to perceive less convergence the less comprehensible and familiar participants judged the non-native speech to be. However, again, it is possible that such effects only arise in an interactive and less predictable setting where comprehension proves a greater challenge. Our results suggest that participants' perceptions of non-native speech may affect the tendency to converge but future studies should try to disentangle the varying influence of these perceptions in different settings.

Studies on phonetic convergence, especially those evaluating the influence of social effects, often include ratings of participants' attitudes to the other speaker along several social dimensions (e.g., Babel, 2010). Here we did not collect ratings of the participants' attitudes towards the speaker nor a measure of their bias towards the model speaker's (assumed) ethnolinguistic group beyond ratings of her accent, so we cannot exclude that these might have also played an additional role, boosting convergence (note that any biases against the non-native speaker would have only diminished the levels of convergence observed here). However, Lewandowski and Nygaard (2018) did not find that their non-native model speakers' lower social status ratings reduced convergence relative to the native speakers in their non-interactive task, nor did Kim (2012) find a role of their participants' IAT scores (attitude towards foreigners) on convergence overall. Nonetheless, many studies have examined the role of bias towards their model speaker's ethnolinguistic group, finding effects even in non-interactive tasks (Abrego-Collier et al., 2011; Babel, 2009). This is something future studies involving non-native speech in particular should consider, in addition to evaluating potential psycho-social effects.

As is common in studies on phonetic convergence (Pardo et al., 2018; Yu et al., 2013), we observed large individual differences, with average perceived convergence rates ranging from 0.33 to 0.77. This large inter-speaker variability has been taken as evidence that speakers may vary in their tendency to phonetically accommodate to other speakers. A recent study even suggests that a speaker's tendency to converge may be a stable trait, demonstrating reliability across time (Wade, Lai, & Tamminga, 2020). Accordingly, recently there have been more studies trying to determine what cognitive individual characteristics underlie individual differences in phonetic convergence (Levitan, 2020; Priva & Sanker, 2019; Weise et al., 2019), as well as examining the relationship between the tendency to phonetically accommodate and other individual traits (e.g., Aguilar et al., 2016; Lehnert-Lehouillier, Terrazas, Sandoval, & Boren, 2020; Lewandowski & Nygaard, 2018; Lewandowski, 2012; Lewandowski & Jilka, 2019; Yu et al., 2013). However, studying individual differences in convergence with natives may be a challenge due to the fact that some participants may sound more similar to the other speaker to begin with. Our results, together with others suggesting that there is convergence to non-native speech, open the door to using non-native varieties to examine individual differences given that there is less chance that, globally, the aligning speech will already be very similar to the target speech. Furthermore, non-native speech could also be particularly useful for studies evaluating the role of acoustic distance and production variability in degree of convergence.

4.2. Acoustic analysis of convergence

Convergence was also evaluated by examining DID measures for multiple acoustic dimensions. The results of these analyses also revealed convergence overall, although the degree and direction varied across items (0.49–0.68) and acoustic dimensions.

Two commonly measured acoustic dimensions in convergence studies, speech rate and f0 (Pardo et al., 2013), were assessed for all items. The model speaker's critical utterances tended to be produced slower and at a higher f0 than most of the participants' at baseline, with the exception of speech rate for *saldo*, likely a product of the model's shortened vowel, and f0 for *schaduw* with its atypical intonation. Participants generally converged to the slower speech rate, except for *saldo*. For that item, participants diverged from the model speaker, producing the word faster than at baseline, probably due to imitation of the reduced vowel. The patterns of convergence to f0 were more complex. The participants seemed to converge to the model's higher values for *stempel*, but for the atypical realization of *schaduw* actually overshot the model's low value to the point of divergence. For *saldo*, which was on the limit of participants' baseline range, participants maintained their baseline values, while convergence to f0 in *atleet* revealed complex interactions with token and ratings of comprehensibility and familiarity.

In addition to f0 and speech rate, the results of the acoustic analyses for *atleet* revealed convergence to stop closure duration, which was also far from participants' mean values. Participants appeared to converge slightly to the model speaker's extended closure of the stop. It is important to note that in this case, the model's closure durations were longer than the participants', rather than shorter, and that the participants converged by extending their closures. This realization is still consistent with voiceless stops, which tend to have longer closures than voiced stops, and thus participants did not move in a direction that would lead to phonetic ambiguity (i.e., atleed).

For *saldo*, the model speaker's final vowel was much shorter than participants', which had greater movement and a more open and back ending sound. However, participants only converged to the duration of the vowel, with a relatively shorter, but still diphthongized vowel.

As regards the word *schaduw*, in addition to adapting to the model speaker's f0, by the end of the task participants had also slightly approximated the model in the more extended duration of the vowel. However, it is worth noting that, once again, this increased duration would not go in the direction of increasing ambiguity with the non-target shorter Dutch [ɑ]. Furthermore, convergence to vowel duration also revealed complex interactions with perceived comprehensibility, accentedness, and familiarity, perhaps due to its atypical intonation.

The acoustic measures of *stempel* reveal that participants adapted to every dimension measured. As can be seen in Table 2, the model speaker's initial consonant had a much lower CoG, like [ʃ], but was not shorter than most participants'

at baseline, as would be expected for [ʃ], which tends to be shorter than [s]. In addition to speech rate and f0 discussed above, participants slightly converged in CoG but diverged from the model speaker's duration of the fricative. It is worth noting that convergence to the model speaker's longer word duration cannot be attributed to adaptation to fricative duration as participants shortened the fricative rather than extended it. Interestingly, although participants diverged from the model in fricative duration, for which the model speaker's value was actually within the participants' baseline range, shortening the fricative could also be interpreted as approximating their own [ʃ] category, an effect which would accompany the lowering of their CoG. The finding of divergence in fricative duration is thus ambiguous, and we can only speculate about whether participants were indeed approximating [ʃ] or diverging from the model speaker. To better study this, an item which was both lower in CoG and shorter in duration than the participants' baseline [s] values would be needed.

Overall, the results of the acoustic analyses show that participants adapted their speech to the non-native stimuli they were exposed to. However, in most cases these changes were very subtle, as is common in phonetic convergence studies (Pardo et al., 2018). Moreover, the degree of convergence varied across items and acoustic dimensions. The finding that degree of convergence varies across items and acoustic dimensions is well-documented in the literature (e.g., Kim & Clayards, 2016; Nielsen, 2011). However, most studies have a large set of items, so any effects of item can be controlled for by, for example, its inclusion as a random effect in a linear mixed effects model. Here we opted for a more restricted but detailed analysis of convergence to our non-native speaker's phonetic patterns, choosing to look at several different phonetic features and acoustic dimensions but represented in only four items, each repeated multiple times. While the restricted item set used here may limit our ability to generalize our precise findings, we can still attempt to extrapolate from the global patterns observed. Overall participants seem to have converged more to speech rate, f0, and even sound duration than to the spectral properties of the vowels. This pattern is consistent with the literature on convergence between native speakers (Pardo et al., 2013, 2017). Moreover, it is worth noting that our participants converged to sound-specific (e.g., vowel and stop closure duration, CoG) in addition to more general acoustic dimensions such as f0 and speech rate. Thus, it does not appear that participants were converging only to the non-linguistic properties of the speaker's voice, but also to her pronunciation of certain sounds, although perhaps less consistently.

The variability in convergence observed here may also have to do with the relative importance of the different acoustic dimensions in distinguishing sounds in Dutch. In that respect, it is worth noting that, while participants did converge to the non-native speaker, in most cases they did so to acoustic dimensions that were less relevant phonetically (e.g., f0, speech rate; duration is a less important cue for distinguishing vowels in Dutch than spectra) or in a direction that would not create phonetic ambiguity (e.g., longer closure for final [t] in *atleet*). This is consistent with Nielsen's (2008, 2011) results where participants converged to extended, but not shortened stop VOT, the latter of which could lead to ambiguity with voiced stops. As Coles-Harris (2017) notes, this suggests a degree of linguistic sensitivity in phonetic convergence. However, it is worth noting that other studies on convergence to non-native speech have observed accommodation patterns that would not preserve phonetic ambiguity.

An exception to this finding in our data is *stempel*, where participants converged to CoG, producing the [s] a little bit more like [ʃ][7]. One possible explanation is that another linguistic factor was acting there, namely: the large production variability of [s] in Dutch, with many native speakers producing the sound more retracted (see footnote 4; Collins & Mees, 2003; Ditewig et al., 2019). Babel (2009) found greater convergence to sounds, such as low vowels, with greater variability in the native population. However, we did not see convergence to the spectral properties of [a:], which also varies a lot regionally (Collins & Mees, 2003). Here it is also worth remembering that [ʃ] production in these contexts is also a salient feature of German-accented Dutch, which, given the large number of Dutch-speaking German students, our participants would be familiar with. Thus, convergence to CoG for this item may also have been subject to the extra influence of social salience and feature awareness, which could have boosted convergence (Walker & Campbell-Kibler, 2015). Furthermore, [ʃ] is illegal in this phonetic context in Dutch, perhaps reducing any potential ambiguity that could result from adapting this sound. Future work can investigate the role of these factors in convergence to non-native speech in a more controlled way.

Some have argued that greater acoustic distance leads to greater convergence (e.g., Babel, 2012). While our study was not set up to test that question and it merits more research, Table 2 nonetheless reveals that participants do not appear to have converged more to dimensions on which the model speaker's values varied more from participants' baseline values. Our results also demonstrate the usefulness for the study of convergence to non-native speech of not only considering participants' original distance from the non-native target, but also considering whether the target is within the native range (similar to the idea of a role for size of phonetic repertoire; Babel, 2009). Future work can address this issue more systematically.

In all, the complex picture revealed by the acoustic analyses reaffirms the need to include multiple acoustic measures as well as acquire the more holistic measurement of perceived convergence (Levitan, 2020; Pardo et al., 2013). Furthermore, the interactions between certain acoustic dimensions and participants' ratings of the non-native speech indicate that complex mechanisms may be at play, which require further study. In this regard, it is interesting to note that, while Kim (2012) did not find that their participants' implicit attitudes towards foreigners were directly linked to their tendency to converge to non-native speakers, their attitudes did interact with their initial acoustic distance to the target speech. Here we did not systematically analyze the role of baseline distance to the model speech, but it is possible that these interacted with our attitude measures and may partly explain the diverse findings for the different acoustic dimensions. Yet another possibility derives

[7] It should be noted that some phoneticians (e.g., Booij, 1999; Collins & Mees, 2003; Mees & Collins, 1982) do not consider the postalveolar fricatives as native Dutch phonemes but rather combinations of [s, z] and [j].

from Lewandowski and Nygaard's (2018) explanation as to why they may have found convergence to non-native speakers while Kim et al. (2011) did not. The authors made reference to the idea that perceptual-motor and socio-communicative mechanisms of convergence may be differentially at play in interactive and non-interactive settings. Our results may go even further to suggest that these different mechanisms may be operating within the same setting and even within the same word, varying per acoustic dimension. This is in line with Levitan's (2020) recent suggestion that each dimension may be differentially susceptible to different convergence mechanisms; while convergence along some acoustic dimensions may interact with attitudes towards the speech or speaker, convergence along other dimensions may be more driven by automatic processes. Here we see just a glimpse of how speakers' perceptions of the target speech on the whole may influence their tendency to converge or diverge from it. More exhaustive studies, in which more precise evaluations of speakers' perceptions are collected and analyzed more robustly, could help to begin to unravel how variations along different dimensions are perceived and accommodated to. The approach used here where convergence along various dimensions was analyzed in conjunction with participants' attitudes to the speech thus highlights a promising avenue to help understand the perplexing patterns of convergence in the literature, as well as the mechanisms underlying convergence to non-native speech.

4.3. Relationship between acoustic and perceptual measures of convergence

The simultaneous use of both perceptual ratings and acoustic measurements enabled us to also assess the relative contribution of different acoustic dimensions to perceived convergence. Our results suggest that raters mainly considered vowel duration, speech rate, and f0 when providing their judgments of convergence. These were the most frequently converged-to dimensions, but also possibly the most salient for raters. Additionally, as mentioned before, intonation contour may have been salient for the raters, although not included in our acoustic dimensions. Furthermore, the negative relationship between convergence to relative closure duration in *atleet* and raters' perception of convergence in the AXB task supports the idea that raters may have occasionally picked up on participants' convergence to dimensions beyond those measured here.

Lewandowski and Nygaard (2018), in their study with native and non-native model speakers, found that convergence on f0 predicted raters' AXB selections for the native models, while vowel spectra were related to AXB ratings for the non-native models. This finding contrasts with our results, but can be partly explained by the fact that participants did not converge to our non-native speaker's vowel spectra. Furthermore, our results are not fully comparable given the different languages, accents, and also the fact that the raters in Lewandowski and Nygaard (2018) study had to judge AXB perceptual similarity to both native and non-native models, while here only a non-native model was present. Our findings for f0 (although inconsistent) and vowel duration are in line with other studies that have found these dimensions to predict AXB ratings (e.g., Pardo et al., 2017). Further research is needed to understand how raters perceive subtle acoustic differences on different dimensions in different accents, as well as what they perceive as non-native realizations. To that end, we agree with Pardo (2013; see also Pardo, Urmanche, Wilman, et al., 2018 that AXB ratings can provide information that escapes the acoustic dimensions chosen, and that more studies should integrate acoustic and perceptual measures of convergence.

4.4. Effectiveness of memory task

In the present study, we developed a novel non-interactive task to study phonetic convergence and demonstrated its effectiveness in eliciting convergence. The advantage of the memory task used here is that it allowed experimental control while providing participants with a credible cover story to keep them naïve as to the purpose of the study. The responses to our debriefing questions reveal that a majority of the participants believed the task was indeed about memory. Of those participants who suspected of the cover story, one of the common reasons cited was the high number of one-item trials and the fact that the task was generally not difficult. Future studies wishing to use a similar task should consider including trials with longer sequences (>3 words) to increase difficulty and perhaps think of a way to include the critical items along with other items in a sequence instead of isolated (keeping in mind findings that greater cognitive load may lead to less entrainment; e.g., Abel & Babel, 2017). In addition, the inclusion of semantically related items or a similar manipulation can be recommended as it seemed to help distract from the true purpose of the study.

Another aspect that led participants to suspect of the memory explanation was precisely the model speaker's non-native accent. We did not offer any explanation for this, but if we had provided participants with a credible reason, likely less would have doubted of the cover story (cf. Walker & Campbell-Kibler, 2015, who suggest increased attention and likelihood of explicit imitation when the model speech varies greatly from the participants' own variety). Similarly, although relatively few participants mentioned the baseline word reading task in their alternative study explanations, providing an adequate explanation for the task, even if just something to the effect of ascertaining knowledge of the words, could aid with the deceit.

Another insight from our debriefing was the high number of participants who, when explicitly asked, recognized imitating the model speaker. Although many participants were not certain if they had engaged in imitation or not, and we did not confirm whether those who admitted imitating effectively did, this raises an interesting question for future studies to explore about the role of imitation awareness in convergence. Similary, here there were not enough participants who had guessed the task was about convergence to properly evaluate its effect on tendency to converge, but it would be interesting to see whether and how such awareness about the study purpose could influence the results.

5. Conclusion

The present study adds to the growing body of evidence that native speakers can display phonetic convergence to

non-native speech, even in a non-interactive situation where socio-communicative motivations should be minimal. Our findings further indicate that degree of convergence to non-native speech is modulated by perceived accentedness, demonstrating the importance of integrating the way speakers perceive the target accent to analyses of convergence to non-native speech.

Most of the previous studies evaluating phonetic convergence to non-native speech have employed interactive tasks, where convergence may be boosted by communicative and social motivations arising from in-person interaction. We are only aware of two other studies that have evaluated convergence to non-native speech with a non-interactive task. Our results add to those findings and extend the observation of convergence to another language combination and to a novel task, one which allows experimental control while maintaining participants' ignorance about the study's goals. Our study adopts a more comprehensive approach, integrating various sounds and acoustic dimensions, both acoustic and perceptual measures, as well as perceptions of the non-native speaker's accent. Overall, our findings have revealed interesting new avenues researchers can explore to further unveil the mechanisms of convergence.

In addition to methodological considerations, our results can also be of use to theories of convergence. Our finding of convergence in a socially impoverished non-interactive task, and despite the lack of a general tendency of native speakers to want to sound like non-native speakers, is hard to integrate with a purely social account of phonetic convergence. However, the observation that perceived accentedness reduced the likelihood that participants be perceived as converging to the model speaker, together with the complex interactions between acoustic measures of convergence and perceptions of the model's accent, are consistent with previous findings that some social factors do still play a role in non-interactive task. Our results therefore are in line with the view that both social and automatic mechanisms underlie convergence, with the relative weight of each depending on the setting in which it is evaluated (e.g., Coles-Harris, 2017) and perhaps even the particular item of analysis. Given the relatively few studies on phonetic convergence to non-native speech, and the fact that our study only included one model speaker while model speakers can vary greatly in their tendency to elicit convergence (Pardo et al., 2017), further research is needed. Nonetheless, our findings demonstrate that native speakers can converge to the speech of a non-native speaker with an unfamiliar accent without any apparent socio-communicative motivations to do so.

CRediT authorship contribution statement

Mónica A. Wagner: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Data curation, Visualization. **Mirjam Broersma:** Conceptualization, Methodology, Writing - review & editing, Supervision. **James M. McQueen:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Sara Dhaene:** Methodology. **Kristin Lemhöfer:** Conceptualization, Methodology, Writing - review & editing, Supervision.

Acknowledgements

The authors would like to express their gratitude to the model speaker for her collaboration. Special thanks to Ronny Bujok for all of his work annotating the audios for this study. We are also grateful to Michael Hess, Victoria Poulton, and Doriana Rinaldi for assistance in processing the audios and to Marjolein van Os and Roeland van Hout for help during stimulus preparation. MAW would further like to acknowledge the Dutch Research Council (NWO) from whom she received support during the writing of this manuscript through a grant (VIDI-276-89-006) awarded to MB.

References

Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech, 60*(3), 479–502. https://doi.org/10.1177/0023830916665652.

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. C. L. (2011). Effects of speaker evaluation on phonetic convergence. *17th International Congress of Phonetic Sciences. ICPhS, 2011*, 192–195.

Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America, 116*(3), 1729. https://doi.org/10.1121/1.1779271.

Aguilar, L., Downey, G., Krauss, R., Pardo, J. S., Lane, S., & Bolger, N. (2016). A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matters. *Journal of Personality, 84*(2), 165–177. https://doi.org/10.1111/jopy.12149.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The Hcrc Map Task Corpus. *Language and Speech, 34*(4), 351–366. https://doi.org/10.1177/002383099103400404.

Baayen, R. H., & Shafaei-Bajestan, E. (2019). languageR: Analyzing Linguistic Data: A Practical Introduction to Statistics (1.5.0) [Computer software]. https://CRAN.R-project.org/package=languageR.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

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. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001.

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.

Babel, M. (2009). Phonetic and social selectivity in speech accommodation [University of California, Berkeley]. https://escholarship.org/content/qt1mb4n1mv/qt1mb4n1mv.pdf.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Bolker, M. B. (2015). Package 'lme4'. *Convergence, 12*(1), 2.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*(2), 145–204.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0. 37. RetrievedFebruary, 3, 2018.

Booij, G. (1999). *The Phonology of Dutch*. Clarendon Press.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3), 255–272. https://doi.org/10.1016/S0167-6393(96)00063-5.

Broos, W. P., Dijkgraaf, A., Van Assche, E., Vander Beken, H., Dirix, N., Lagrou, E., et al. (2019). Is there adaptation of speech production after speech perception in bilingual interaction? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(7), 1252.

Brown, G., Anderson, A., Yule, G., & Shillcock, R. (1983). *Teaching Talk*. Cambridge University Press.

Coles-Harris, E. H. (2017). Perspectives on the motivations for phonetic convergence. *Language and Linguistics Compass, 11*(12). https://doi.org/10.1111/lnc3.12268 e12268.

Collins, B., & Mees, I. M. (2003). The phonetics of English and Dutch. Brill.

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes, 23*(4), 528–556. https://doi.org/10.1080/01690960801920545.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication, 30*(2–3), 109–119. https://doi.org/10.1016/S0167-6393(99)00040-0.

Delvaux, V., & Soquet, A. (2007). *The influence of ambient speech on adult speech productions through unintentional imitation., 64*(2–3), 146–173.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In Advances in experimental social

psychology (Vol. 33, pp. 1–40). Academic Press. https://doi.org/10.1016/S0065-2601(01)80003-4.
Ditewig, S., Pinget, A.-F., & Heeren, W. (2019, September). Regional variation in the pronunciation of /s/ in the Dutch language area [Text]. https://doi.org/info:doi/10.5117/NEDTAA2019.2.003.DITE.
Felker, E., Troncoso-Ruiz, A., Ernestus, M., & Broersma, M. (2018). The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input. *The Journal of the Acoustical Society of America, 144*(4), EL304–EL309. https://doi.org/10.1121/1.5063809.
Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics, 14*(1), 3–28. https://doi.org/10.1016/S0095-4470(19)30607-2.
Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*(3), 396–413. https://doi.org/10.1016/S0749-596X(03)00072-X.
Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology, 4*, 340.
Gasiorek, J., Giles, H., & Soliz, J. (2015). Accommodating new vistas. *Language & Communication, 41*, 1–5.
Giles, H., & Ogay, T. (2007). Communication accommodation theory. In Explaining communication: contemporary theories and exemplars. Mahwah, NJ: Lawrence Erlbaum (Whaley, B. B., Samter, W.).
Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In Contexts of accommodation: Developments in applied sociolinguistics (Vol. 1). Cambridge University Press.
Giles, H. (1973). Accent mobility: A model and some data. Anthropological Linguistics, 15(2), 87–105. JSTOR.
Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review, 14*(2), 214–237. https://doi.org/10.1177/1088868309359288.
Goblirsch, K. (2015). Language contact and consonant shift in Germanic: The witness of aspiration. In J. O. Askedal & H. F. Nielsen (Eds.), *early Germanic languages in contact*. John Benjamins Publishing Company.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251.
Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11*(4), 716–722.
Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association, 32*(2), 141–174. https://doi.org/10.1017/S0025100302001020.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464.
Gussenhoven, C., & Broeders, A. (1997). *English pronunciation for student teachers* ((2nd ed.).). Wolters Noordhoff.
Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language, 81*, 72–90. https://doi.org/10.1016/j.jml.2015.01.001.
Ivanova, I., Costa, A., Pickering, M. J., & Branigan, H. P. (2007). Lexical alignment of L1 speakers with L2 speakers. AMLaP-2007 (Architectures and Mechanisms of Language Processing Conference), Türkü, Finland.
Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007.
Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643.
Kim, D., & Clayards, M. (2016). Individual differences in the relation between perception and production and the mechanisms of phonetic imitation 3113–3113. *The Journal of the Acoustical Society of America, 140*(4). https://doi.org/10.1121/1.4969741.
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.
Kim, M. (2012). Phonetic accommodation after auditory exposure to native and nonnative speech.
Kuijpers, C. T. L. (1993). Temporal aspects of the voiced–voiceless distinction in speech development of young Dutch children. *Journal of Phonetics, 21*(3), 313–327. https://doi.org/10.1016/S0095-4470(19)31341-5.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13).
Lehnert-Lehouillier, H., Terrazas, S., Sandoval, S., & Boren, R. (2020). The Relationship between Prosodic Ability and Conversational Prosodic Entrainment. In Speech prosody (Urbana, Ill.) (Vol. 2020). https://doi.org/10.21437/SpeechProsody.2020-157.
Levitan, R. (2020). Developing an integrated model of speech entrainment. In *Proceedings of the twenty-ninth international joint conference on artificial intelligence* (pp. 5159–5163).
Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Twelfth annual conference of the international speech communication association*.
Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality & attention. *Frontiers in Communication, 4*, 18.
Lewandowski, E. M., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America, 144*(2), 620–633. https://doi.org/10.1121/1.5038567.
Lewandowski, N. (2012). Talent in nonnative phonetic convergence. http://elib.uni-stuttgart.de/handle/11682/2875.
Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6.
Marcoux, K. P., & Ernestus, M. T. C. (2019). Pitch in native and non-native Lombard speech. In *Proceedings of the 19th international congress of phonetic sciences* (pp. 2605–2609).
McCloy, D. R., & McCloy, M. D. R. (2016). Package 'phonR'. *Sigma, 2*, 2.
Mees, I., & Collins, B. (1982). A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetic Association, 12*(1), 2–12. https://doi.org/10.1017/S0025100300002358.
Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285–310.
Natale, M. (1975). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills, 40*(3), 827–830. https://doi.org/10.2466/pms.1975.40.3.827.
Nelson, L. R., Signorella, M. L., & Botti, K. G. (2016). Accent, gender, and perceived competence. *Hispanic Journal of Behavioral Sciences, 38*(2), 166–185. https://doi.org/10.1177/0739986316632319.
Nielsen, K. Y. (2008). *Word-level and feature-level effects in phonetic imitation*. Los Angeles: University of California. http://search.proquest.com/openview/91d6a85bb6b34c5a1e969c3550eaaccf/1?pq-origsite=gscholar&cbl=18750&diss=y.
Nielsen, K. Y. (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. (2013). Measuring phonetic convergence in speech production. *Cognitive Science, 4*, 559. https://doi.org/10.3389/fpsyg.2013.00559.
Pardo, J. S. (2016). Catching the drift: Carol A. Fowler on phonetic variation and imitation. *Ecological Psychology, 28*(3), 171–175. https://doi.org/10.1080/10407413.2016.1195190.
Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183–195. https://doi.org/10.1016/j.jml.2013.06.002.
Pardo, J. S., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., et al. (2018). The Montclair Map Task: Balance, efficacy, and efficiency in conversational interaction 0023830918775435. *Language and Speech*. https://doi.org/10.1177/0023830918775435.
Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics, 79*(2), 637–659.
Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics, 69*, 1–11. https://doi.org/10.1016/j.wocn.2018.04.001.
Pardo, J. S. (2010). Expressing oneself in conversational interaction. In Expressing Oneself / Expressing One's Self: Communication, Cognition, Language, and Identity (E. Morsella, pp. 183–196). Psychology Press.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056.
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495.
Priva, U. C., & Sanker, C. (2019). Limitations of difference-in-difference for measuring convergence. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 10*(1), 15. https://doi.org/10.5334/labphon.200.
R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.
Rao, G. N. (2013). Measuring phonetic convergence: Segmental and suprasegmental speech adaptations during native and non-native talker interactions. https://repositories.lib.utexas.edu/handle/2152/23105.
Rietveld, A. C., & van Heuven, V. J. (2009). Algemene fonetiek. *Coutinho*.
Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*(4), 421–436. https://doi.org/10.1006/jpho.1997.0051.
Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Accommodation theory 25 years on. In The new handbook of language and social psychology (pp. 33–56). John Wiley and Sons.
Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I. *Language and Speech, 12*(2), 80–102.
Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society, 37*(2), 241–254.
Vallabha, G. K., & Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *The Journal of the Acoustical Society of America, 116*(2), 1184–1197. https://doi.org/10.1121/1.1764832.
van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics, 32*(4), 455–491. https://doi.org/10.1016/j.wocn.2004.05.001.

van der Harst, S. (2011). The vowel space paradox: A sociophonetic study on Dutch [Radboud University Nijmegen]. https://www.lotpublications.nl/the-vowel-space-paradox-the-vowel-space-paradox-a-sociophonetic-study-on-dutch.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(Pt 4), 510–540.

Wade, L., Lai, W., & Tamminga, M. (2020). The reliability of individual differences in VOT imitation 0023830920947769. *Language and Speech*. https://doi.org/10.1177/0023830920947769.

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

Weise, A., Levitan, S. I., Hirschberg, J., & Levitan, R. (2019). Individual differences in acoustic-prosodic entrainment in spoken dialogue. *Speech Communication, 115*, 78–87. https://doi.org/10.1016/j.specom.2019.10.007.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics, 75*(3), 537–556. https://doi.org/10.3758/s13414-012-0404-y.

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution, 3*(1), 129–137.

Wooldridge, B. (2001). 'Foreigner Talk': An important element in cross-cultural management education and training. *International Review of Administrative Sciences, 67*(4), 621–634. https://doi.org/10.1177/0020852301674002.

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

Zeileis, A., Meyer, D., & Hornik, K. (2007). Residual-based shadings for visualizing (conditional) independence. *Journal of Computational and Graphical Statistics, 16*(3), 507–525.