Journal of Phonetics 108 (2025) 101376

Contents lists available at ScienceDirect

Journal of Phonetics

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

# Individual uniformity in phonetic imitation: Assessing the stability of individual variability across features and tasks ☆

Jessamyn Schertz *

*Department of Language Studies, University of Toronto Mississauga, Canada*

ARTICLE INFO

*Article history:*
Received 11 July 2024
Received in revised form 7 November 2024
Accepted 16 November 2024
Available online 2 December 2024

*Keywords*:
Phonetic imitation
Phonetic convergence
Individual differences

ABSTRACT

Extensive individual variability has been reported in both spontaneous phonetic convergence and in explicit phonetic imitation tasks. This work tests the consistency of these individual patterns: are some individuals just globally more imitative, showing greater-than-average imitation regardless of the specific phone being imitated, and regardless of the type of imitation, or does an individual's extent of imitation depend heavily on the phonetic content or the type of task? We examine the stability of individual variability in imitation of two types of subphonemic differences (VOT of voiceless stops and F2 of the vowels /æ/ and /u/), in two types of imitation tasks (implicit and explicit). We found that individuals' degree of imitation was significantly related across different phones within the same class (e.g., imitation of /p/ vs. /t/) in both implicit and explicit imitation tasks, but that individuals' degree of imitation of phones from different classes (e.g., imitation of stops vs. vowels) was only related in explicit, but not implicit, imitation. Findings are consistent with the idea that general cognitive or personality traits may govern individual variability in explicit imitation, but challenge the idea that they play any measurable role in predicting individual variability in implicit or spontaneous imitation. We also found a weak but significant correspondence between individual performance on the implicit and explicit imitation tasks, providing evidence that the two tasks rely on shared mechanisms, as well as a significant relationship between discrimination performance and explicit, but not implicit, imitation.



## 1. Introduction

"Phonetic imitation," defined broadly, can refer to any situation in which an individual's speech becomes more closely aligned with the phonetic properties of speech they are exposed to. Phonetic imitation occurs in diverse contexts and at different levels of active awareness, from "phonetic convergence," where conversation partners may unconsciously start to sound more like one another, to metalinguistic tasks like imitating the idiosyncrasies of a specific accent (e.g., Pardo, 2013; Myers et al., 2024). Previous work has documented considerable variability in the extent of imitation, in that some individuals show much more imitation than others (e.g. Wade et al., 2021). However, the scope and generalizability of these individual patterns of imitation is an open question: are some individuals simply more global imitators, regardless of the context of imitation or the nature of the exposure speech, or does the extent of an individual's imitation ability depend heavily on the type of task, and/or on the specific phonetic characteristics being imitated? An understanding of the uniformity of individual variability in performance on imitation tasks is important for understanding the cognitive architecture underlying (different types of) phonetic imitation and is a necessary step in identifying the "individual differences," i.e. systematic linguistic, cognitive, and/or social characteristics intrinsic to individuals, that predict how "imitative" they are.

The primary goal of this work is to examine the consistency of individuals' patterns of imitation across two different phonetic features (voice onset time (VOT) in voiceless stops and F2 in vowels), in two different types of imitation (implicit and explicit). If general cognitive or personality-based factors are

☆ Supplementary materials, including all stimuli, data, and analysis code used in this work are available here or by contacting the author.

* Address: Department of Language Studies, University of Toronto Mississauga, Maanjiwe nendamowinan, 4th floor, 3359 Mississauga Road North, Mississauga, ON L5L1C6 Canada.
*E-mail address:* jessamyn.schertz@utoronto.ca

https://doi.org/10.1016/j.wocn.2024.101376
0095-4470/

strong drivers of individual differences in imitation, we expect to see a strong relationship between a given speaker's imitation of different features; for example, someone who shows greater-than-average imitation of stops should also show greater-than-average imitation of vowels. Along with testing stability of imitation across features in each type of task, we also test the relationship between individual performance across the implicit vs. explicit tasks, and, via a discrimination task, assess the extent to which individual variability in imitation can be attributed to differences in perception of the relevant features.

### 1.1. Implicit vs. explicit phonetic imitation tasks

Phonetic imitation has been examined across a wide range of contexts that vary in the level of interaction, the nature of the exposure speech, and the instructions given to participants. This work examines the structure of individual variability in implicit and explicit imitation tasks that differ in whether or not they are designed to elicit active, conscious imitation by the speaker. Implicit imitation (also called *spontaneous imitation* or *phonetic convergence)* encompasses any task designed to elicit phonetic imitation without conscious intent to imitate, including that found in naturalistic interactions as well as in non-interactive, word repetition tasks (e.g. Goldinger, 1998; see discussion and examples in Pardo, 2013). Explicit imitation (or *forced imitation*) encompasses any situation where the stated goal of the task is for participants to sound like the speech they hear, whether it is imitation of individual words (e.g. Pinget, 2022) or mimicking an unfamiliar accent (e.g. Mora et al., 2014; Kopečková et al., 2021; Myers et al., 2024).

Implicit and explicit imitation have generally been examined in independent lines of research, but a small set of studies has directly compared the two types. These studies have for the most part found greater degrees of imitation in explicit than in implicit tasks (Dufour & Nguyen, 2013 for F1 of vowels; Garnier et al., 2013 for f0; Sato et al., 2013 for f0 and vowel F1; Clopper & Dossey, 2020 for perceptual assessment of imitation of monosyllabic words; an exception to this pattern is Pardo et al., 2010, who found no increase in convergence in with explicit instructions to imitate in a conversational task). While the precise nature of the relationship between them is an open question, implicit and explicit phonetic imitation have been proposed to be distinct processes that nevertheless draw on shared mechanisms (Dufour & Nguyen, 2013; Schertz et al., 2023). There is good evidence that there is an automatic component to imitation, proposed to be a direct consequence of interaction between incoming speech and listeners' representations (e.g. Goldinger, 1998; Pickering & Garrod, 2013), but which may also be modulated by linguistic or social factors, even when below the level of conscious awareness (see Coles-Harris, 2017, for discussion). Any such automatic component would be expected to be shared by, and play a role in, both implicit and explicit imitation.

At the same time, implicit and explicit tasks likely differ substantially in their "targets" of imitation, in the mechanisms involved in the process of imitation, and in the factors that might influence the extent of imitation. For explicit imitation, the intended target of imitation is faithful replication of relevant properties of the incoming speech.[1] The intended target for implicit imitation is less transparent, but given the absence of instructions to imitate, it is likely that participants assume that faithful imitation is *not* the intended target, particularly when the stated goal of the task is something other than imitation (e.g., when it is presented to the participants as a memory task, as in Delvaux & Soquet, 2007, and as in the current work). Work by Zellou et al. (2017) suggests that the targets for implicit imitation are influenced both by the immediately preceding input (i.e. a word that has just been heard) and by the longer-term average properties of a model talker's productions.

In terms of the mechanisms invoked during imitation, explicit tasks direct attention to the phonetic characteristics of the incoming speech, prompting participants to identify what they consider to be its salient properties and to manipulate their production patterns to replicate those properties, whereas implicit tasks do not. In one sense, therefore, explicit imitation tasks recruit additional processes not necessarily invoked in implicit imitation tasks (Dufour & Nguyen, 2013; Schertz et al., 2023). In another sense, however, because of the absence of direct instruction, "successful" imitation in an implicit task requires engagement from the participant in ways that are not strictly necessary in an explicit task. Baseline abilities to perceive and produce the relevant target features are necessary for both types of tasks; however, for imitation to occur in an implicit task, the participant must additionally track and attend to the relevant dimension without prompting. Furthermore, they need to determine (not necessarily consciously) that at least some degree of convergence or imitation is the intended goal, since the target is not explicitly specified in the instructions.

The primary goal of this work is to examine the extent of individual uniformity in imitation of phonetic features, i.e. whether some individuals tend to be globally "more imitative" than others, or whether imitation is feature-specific. However, given the considerable differences between the goals and processes associated with implicit and explicit imitation, the degree of feature-specificity could depend on the type of imitation being tested, as individual variability on the two types of tasks may be governed by different factors (as will be discussed below). Therefore, we test the extent of uniformity across features in two different lab-based word repetition tasks in which participants hear words and are asked to repeat them, either without or with explicit instructions to imitate. In addition, we test individual uniformity across the two tasks (i.e. whether individuals who show greater-than-average imitation in explicit tasks are the same ones who do so in implicit tasks), as this may be informative about the cognitive architecture of explicit vs. implicit imitation, and the relationship between the two.

### 1.2. Predictors of variability in phonetic imitation

Phonetic imitation is fundamental to broader domains such as sound change, language learning, and conversational inter-

[1] Which properties are "relevant" are not static and may vary based on the properties of the participants or the task instructions. For example, consider two sentence repetitions tasks that are identical except for the task instructions: one asks participants to imitate the *voice* they heard, and the other asks participants to imitate the *accent* they heard. Even if the material to be imitated is identical, these different instructions might elicit attention to, and imitation of, different phonetic properties.

action, and as such, there has been considerable interest in trying to identify the factors that are correlated with variability in imitation, as well as the underlying causes of this variability. The extent of imitation can differ based on the linguistic or social properties of the exposure speech on a group level, and individual characteristics of the imitators themselves have been proposed to further facilitate/inhibit imitation. This section briefly outlines some examples of group-level phonetic and social selectivity before turning to reported predictors of individual variability in phonetic imitation.

Previous work has reported phonetic selectivity in imitation, i.e., that the extent of imitation may vary systematically across different sounds or classes of sounds. Some of these differences have been ascribed to the phonological status of the target sound or feature. For example, Mitterer and Ernestus (2008) found that phonetic detail was more likely to be imitated when it was phonologically relevant: Dutch speakers faithfully imitated stops differing in the presence vs. absence of prevoicing, but did not imitate differences in the duration of prevoicing. This discrepancy was attributed to the fact that the presence of voicing is the primary differentiator of the Dutch laryngeal contrast (e.g. /b/ vs. /p/), whereas the duration of prevoicing is not relevant to the contrast (see also Olmstead et al., 2013). Nielsen (2011) found that English speakers produced longer VOTs in phonologically voiceless stops (e.g. /p/) after exposure to stimuli with lengthened VOTs, but showed no analogous modifications when exposed to shortened VOTs. Nielsen (2011) noted that there were multiple possible explanations for this asymmetry, including a production-based consideration for contrast maintenance, i.e., that producing shortened VOT might impinge on a different phonological category, and a perception-based explanation, i.e., that the tokens with lengthened VOT were more perceptually salient than those with shortened VOT, supported by subsequent work (Nielsen & Scarborough, 2015).

Social factors also influence the extent of imitation. Positive attitudes toward a talker have been reported to facilitate imitation overall (e.g., Babel, 2012; Yu et al., 2013; Myers et al., 2024), and the extent of imitation has also been shown to depend on sociophonetic properties of specific sounds. Clopper and Dossey (2020) tested imitation of a non-prestige variety of English (Southern American English) which differed from Standard American English in several features (/aɪ/ monophthongization, back vowel fronting, and word duration). They found that in implicit imitation, participants avoided convergence to a socially stereotyped sound variant (/aɪ/ monophthongization), even though they did converge to other systematic phonetic variation in the same model talker. This provided strong evidence that metalinguistic awareness or stereotypicality of a feature, and not simply its presence in a non-prestige variety, can modulate imitation. Similarly, Lee-Kim and Chou (2024) found that female speakers of Taiwanese Mandarin avoided convergence towards a socially stigmatized variant (highly retroflexed sibilants), despite showing imitation more generally. However, the extent of social selectivity may vary depending on the nature of the task: the avoidance of convergence to /aɪ/ monophthongization found in Clopper and Dossey (2020) was somewhat mitigated for listeners who were given explicit instructions to imitate.

The studies discussed in this section so far have examined how properties of the exposure speech, whether it be phonetic characteristics or social connotations, affect the extent of imitation. At the same time, there has been a growing interest in identifying individual-level predictors, i.e. how social, cognitive, or linguistic characteristics inherent to individuals may influence imitation (see review in Myers et al., 2024). Existing work has found that personality traits like the Big Five dimensions of Openness and Neuroticism (Yu et al., 2013; Lewandowski & Jilka, 2019), Attention-Switching, a subcomponent of the Autism Spectrum Quotient (Yu et al., 2013), and rejection sensitivity (Aguilar et al., 2016) correlate with individual variability in spontaneous imitation, and Myers et al. (2024) found that individuals' openness and musical perception correlated positively with explicit imitation of different regional accents of English (see also Coumel et al., 2019, who found that musical talent correlates with imitation of foreign speech). At the same time, it is worth noting that while many correlations have been reported, findings are not always consistent across studies. The provenance of these cross-study discrepancies is not clear, but some potential reasons are spurious results, lack of power, or methodological/task-based differences across studies: as discussed by Myers et al. (2024), individual-level predictors may apply differently to different tasks.

Along with the considerations of phonetic selectivity discussed above, linguistic factors, both production- and perception-based, have also been shown to modulate imitation on an individual level. In terms of production, articulatory ability or flexibility has been reported to correspond with more phonetic convergence in a conversational task (Lee et al., 2021), in explicit imitation of regional accents (Myers et al., 2024), and in explicit imitation of foreign speech (Reiterer et al., 2013), and bilingualism has been reported to facilitate better imitation of novel accents (Spinu et al., 2018; Spinu et al., 2020). The relative position of an individual in an ongoing sound change has also been found to correspond with imitative ability: Pinget (2022) showed that the extent of imitation of a phonetic feature currently undergoing change (devoicing of Dutch /v/) differed systematically based on the extent to which an individual participated in the change (see also Zellou & Brotherton, 2021).

In terms of perception, there is growing evidence of systematic individual differences in listeners' perceptual sensitivity or gradience in phonetic perception (e.g. Kapnoula et al., 2017), relative reliance on phonetic dimensions (see Schertz & Clare, 2020, for review), and the structure of their phonetic/phonological representations more generally (Yu & Zellou, 2019). Since imitation relies on perception, these differences in perceptual strategies or representations might be expected to contribute to the individual variability found in performance on imitation tasks. Direct tests of the correspondence between individual-level perception and imitation have provided some evidence for this link, but not in an entirely straightforward way. Kim & Clayards (2019) examined explicit imitation of spectral and durational cues to the English /ɛ/-/æ/ contrast and perceptual weighting of the same dimensions. While it might be expected that higher perceptual weighting of a dimension would relate to more imitation of the same dimension, this was not the case. Instead, individual reliance on spectral cues

in perception was related to more faithful imitation of durational (but not spectral) variability, while individual reliance on durational cues was not predictive of imitation on either dimension. The authors concluded that general perceptual acuity, rather than attention to individual dimensions, was likely to be the driver of individual variability in imitation, given that the only significant correlation between perception and imitation occurred across different dimensions (see also Hazan & Rosen, 1991; Clayards 2018).

### 1.3. Structured variability in phonetic imitation

The existing empirical findings and the proposals about the cognitive architecture of different types of imitation set the groundwork for some general predictions about expected individual consistency across tasks and features.

#### 1.3.1. Expectations for task-based consistency

Previous work has proposed that implicit and explicit imitation rely on shared resources, including some degree of automaticity, but that the two tasks also recruit distinct processes (Dufour & Nguyen, 2013; Schertz et al., 2023). If this is correct, we might expect to see a relationship between individual performance on the two tasks, since any variability linked to the automatic component should influence performance on both tasks. If both tasks draw heavily on this shared automatic component, this relationship should be relatively strong. At the same time, as discussed above, the two tasks differ in many ways, including the targets of imitation, the perceptual/attentional processes necessary for imitation, and the effect of social factors on imitation in each task. While not an exhaustive list, here we lay out three individual-level predictors that could potentially affect the two tasks in different ways, leading to a disconnect in individual performance between implicit and explicit imitation.

First, some individuals might be more willing or motivated than others to consciously try to attain the stated goal of a task. We expect that this factor would play a stronger role in explicit than in implicit imitation: since the intended target for this task is faithful imitation, then all else being equal, an individual who is highly motivated to try to "succeed" in the task will show more imitation than one who is not. On the other hand, for the implicit task, since the stated target is not faithful imitation, we would not necessarily expect differential performance by these same two individuals.

Second, individual differences in phonetic cue weighting, or the relative importance of different dimensions in phonetic representations (e.g., Schertz & Clare, 2020) might differentially impact imitation on the two types of tasks. Previous work has shown that patterns of imitation are governed not just by the phonetic characteristics of the exposure speech itself, but also by the phonetic and phonological properties of the relevant sounds on a more abstract level (Mitterer & Ernestus, 2008; Kwon 2019). While both explicit and implicit imitation require baseline sensitivity to the relevant differences, implicit imitation additionally requires the participant to pay attention to the relevant dimension, even when not explicitly directed to do so. It is therefore possible that individual differences in sensitivity and cue weighting may be a stronger predictor of individual performance in implicit than in explicit imitation tasks.

Third, while social factors have been shown to play a role in both implicit and explicit imitation (implicit: Babel, 2012; Yu et al., 2013; Lee-Kim & Chou, 2024; explicit: Myers et al., 2024), they could potentially affect the two tasks differently. In a direct comparison of the two tasks, Clopper and Dossey (2020) found that participants' avoidance of imitation of a socially stigmatized phonetic variant in an implicit word repetition task was somewhat mitigated when participants were explicitly instructed to imitate. Based on this finding, individual differences in attitudes toward the target talker or perceived social connotations of the target speech may be expected to play more of a role in implicit than in explicit imitation.

#### 1.3.2. Expectations for feature-based consistency

The various individual predictors discussed above might be expected to apply at different levels of generality, operating over different domains of linguistic structure. As discussed above, non-language-specific predictors (e.g. personality or cognitive traits) would generally be expected to apply equally to imitation of all features. Furthermore, some linguistic predictors, like perceptual acuity (Kim & Clayards, 2019) and general articulatory flexibility (Reiterer et al., 2013) have been proposed to apply at a general, rather than (or in addition to) at a phone- or feature-specific level. If any of these general characteristics play a large role in predicting individual imitative ability, we should see a correspondence in individual patterns across features; i.e., those individuals who imitate stops more should also imitate vowels more.

At the same time, some of the predictors discussed would be expected to apply to specific sounds, with the effect potentially differing across individuals. For example, if imitation is inhibited by negative stereotypes of a specific variant (as in Clopper and Dossey's (2020) finding of avoidance of convergence to monophthongized /aɪ/), and if listeners differ in the strength of its associated social connotations, then this would affect the relative ranking of individuals' imitation of /aɪ/ in a way that would not correspond with the relative ranking of imitation of a different vowel.

Finally, some individual predictors might be expected to operate not just on specific phones, but more broadly on phonetic classes. Individual variability in speech production has been shown not to be random, but instead governed by *target uniformity* (Chodroff & Wilson, 2017, 2022): the idea that talker-specific production norms can apply to phonetic or phonological classes, such that an individual talker will show similar realizations of a feature across all sounds in that class (e.g., a talker who shows longer-than-average VOT for /p/ is also likely to show longer-than-average VOT for /t/). While target uniformity in previous work has generally been considered in terms of a talker's baseline production patterns, we similarly might expect that the extent to which they imitate a given feature also varies in a structured way, based on the substantial evidence for individual differences in speech perception more generally (as reviewed in Yu & Zellou, 2019). For example, some listeners might attend differentially to certain types of information (e.g. one listener might have strong perceptual acuity for spectral characteristics like formants, but less for temporal characteristics like VOT, see Makashay, 2003), and/or listeners might differ in the relative importance of a given dimension in their phonetic representations (as reviewed in Schertz & Clare, 2020).

In sum, observing the domain of individual uniformity in imitation may help to inform about which types of factors play meaningful roles in predicting the extent of imitation. If only non-phonetically-selective factors play a role, we would expect individual patterns to be stable across different phones and features. If only phonetically-selective factors play a role, then we would expect a complete disconnect between individual performance across different phones. In reality, of course, multiple predictors likely contribute to variability in imitation. Assuming that this is the case, we would expect to see *some* correspondence in individual patterns across the board, regardless of what phone is being targeted, since some of these factors are non-phonetically-selective. At the same time, because of the simultaneous presence of factors that *are* phonetically selective, we would expect even stronger correspondences for phones sharing a phonetic feature. We would expect the strongest correspondences for imitation of the same phone.

Finally, as discussed above, individual predictors may play different roles in implicit vs. explicit imitation tasks. Therefore, depending on the relative roles of phonetically-selective vs. non-selective predictors in the two tasks, we expect that the degree of individual uniformity across features may differ between for implicit and explicit imitation. We do not have strong a priori predictions for what these differences might be, but in interpreting the results, we will consider what any disconnect in individual uniformity across implicit vs. explicit imitation can tell us about the roles of different predictors.

#### *1.3.3. Previous work examining individual uniformity in imitation*

The previous sections laid out a catalogue of characteristics reported to influence individual imitation performance. These studies encompass a diverse range of tasks, including both implicit and explicit imitation, and quantify imitation using different perceptual and/or acoustic metrics. The claims are generally tested by examining how a proposed predictor relates to individuals' performance on a specific task, with significant correlations between the predictor and individual performance taken as support for that predictor. For example, Yu et al. (2013) reported that individuals with greater Openness scores in the Big Five personality test showed more implicit imitation of VOT. The authors suggested that this finding might be attributable to these individuals having a higher level engagement with the exposure materials, leading them to pay more attention to the exposure speech and show more imitation overall. This sort of interpretation relies on two assumptions. First, in order to have confidence in the validity of the specific correlation (e.g. Openness and implicit VOT imitation), the individual indices of imitation collected in the study must be a reliable metric of the individual's performance in the specific task. Second, for this finding to be taken as evidence for Openness influencing imitation more generally, it must be the case that the individual indices are also more generally representative of how much they would imitate other features, or how much imitation they would show on other tasks.

Wade et al. (2021) tested the first assumption, stability of individual variability across multiple sessions of the same task. They found that variability in individuals' performance in imitation of lengthened VOT in an implicit word repetition task was stable across different sessions of the same task on different days, confirming "narrow reliability" of the individual indices of imitation for that task. However, there has been little work examining the stability of individual variability across tasks and across features, and the studies that do exist have not shown evidence of a high degree of consistency.

There is very little work that directly examines the degree of individual uniformity across the two types of tasks, i.e., whether those individuals who show more spontaneous imitation than average also imitate more in an explicit task. Pinget (2022) tested both implicit and explicit imitation of devoicing of Dutch /v/, a feature undergoing change, and found that leaders of sound change (those who were further advanced in the sound change in their baseline productions) showed *more* imitation in the implicit task, but *less* imitation in the explicit task. While this suggests a disconnect between individual performance in the two types of imitation, this interpretation is complicated by the fact that the targets of imitation (i.e. the exposure stimuli) in the two tasks were different (see also Pardo et al., 2018 for a disconnect in individual performance across shadowing and conversational tasks). The only work we are aware of that has directly examined the relationship between individual performance across implicit vs. explicit imitation of the same exposure speech (Sato et al., 2013) found no correspondence in a small subset of participants who had completed both tasks (n = 12); however, the modest number of participants makes it difficult to draw strong conclusions about this lack of relationship. It is therefore an open question whether there is a detectable correspondence between an individual's tendency for spontaneous convergence and how much they imitate when explicitly instructed to do so.

Turning to the question of cross-feature uniformity: several studies have examined whether individuals are stable in their imitation of different features within the same task. All of the studies we are aware of have failed to find any evidence of consistency across features. Cohen Priva and Sanker (2020) measured convergence in six linguistic characteristics (f0 median, f0 range, speech rate, and three lexical/sentential factors) in a corpus of conversational speech and found no correspondence between individual rates of convergence across features. Similarly, Sanker (2015) found no individual consistency in convergence of eight acoustic characteristics (including formant measures, f0, intensity, and turn-transition timing) in a conversational task, and similar null results were found in conversational speech in work by Weise & Levitan (2018). However, it is worth noting that all of these studies examined spontaneous convergence in conversational interaction. It is therefore possible that there may indeed be individual uniformity in imitation across features, but that the relationship is obscured by the extensive variability inherent to conversational tasks.

In sum, the question of consistency of imitation across tasks and features is not well-established: there is no positive evidence for consistency, but there have been few direct tests, and those that exist may not have the best chance of finding an effect, either due to small sample size or the presence of substantial other sources of variability. The current study was designed to test of whether individual stability in imitation across features and tasks would be detectable in a more controlled paradigm. Knowing the extent of uniformity of individual patterns is necessary to gauge how confident we can be in

generalizing findings of one study to others. In addition, examining the structure of this individual variability may be informative about the processes governing explicit vs. implicit imitation, the types of factors that play substantial roles in predicting each type of imitation, and at what level of generality they apply.

### 1.4. Current study

Establishing the presence and nature of individual uniformity in phonetic imitation is a necessary step in identifying the most reliable types of predictors of individual differences in imitation, and how these predictors may apply differently depending on the type of imitation. In addition, from a practical point of view, it provides an idea of how reliable results from a single task are as a proxy for an individual's proclivity for imitation in other features and/or tasks. In this work, we test the extent to which individuals show stable levels of phonetic imitation across features (VOT of voiceless stops /p/ and /t/ and F2 of vowels /æ/ and /u/) in two types of imitation tasks: implicit word repetition and explicit imitation. Previous work has found both implicit and explicit imitation of both VOT and vowel formants. VOT is likely the best-studied dimension, and a large number of studies have found that speakers adjust their VOT when exposed to speech with artificially-lengthened VOT (e.g. Shockley et al., 2004; Nielsen, 2011; Yu et al., 2013; Wade et al., 2021 for implicit imitation; Schertz & Paquette-Smith, 2023 for explicit imitation). Imitation of vowel formant values has also been found in both implicit and explicit imitation (e.g., Tilsen 2009; Babel, 2012; Pardo, 2017; Dufour & Nguyen, 2013 for implicit; Dufour & Nguyen, 2013 for explicit), but there have also been cases where imitation was not found: notably, Uluşahin et al., 2023 found no imitation of acoustically large but subphonemic differences in F2 of Dutch vowels.

We begin with group-level analyses testing the presence and extent of imitation of each feature in the two tasks, as well as the discriminability of the target differences. We then turn to the test of stability of individual variability in imitation performance. For each type of imitation, we first test **whether an individual's extent of imitation is stable across the different classes of phones** (i.e. whether an individual who shows greater-than-average imitation of stops also shows greater-than-average imitation of vowels), and **whether there is greater stability across phones within a phonetic class**. If general cognitive or personality traits strongly influence spontaneous convergence, we would expect individual patterns to be correlated across all phones in both tasks. If there are general cognitive or personality traits that would be expected to apply exclusively in the explicit imitation task (e.g. task motivation, willingness to diverge from production norms), we would expect individual patterns to be more strongly correlated across phones in explicit imitation than in implicit imitation. Finally, to the extent that imitation is influenced by phonetic factors (either perception- or production-based) that might be feature- or phone-specific, we expect that within-class correspondences will be stronger than across-class correspondences.

Along with testing the stability of cross-phone correspondences, we also examine two additional questions. First, we examine **the relationship between individuals' implicit and explicit imitation**, testing whether those participants who show a high degree of explicit imitation are also those who show greater-than-average spontaneous imitation, or whether performance on these two tasks appears to be independent. Finally, we test **the relationship between individuals' extent of imitation and their discrimination accuracy** in order to assess the extent to which individual variability in imitation might be attributable to perceptual factors.

## 2. Methods

### 2.1. Participants

Data from 133 participants (63 women, 69 men, 1 nonbinary, aged 18–41 at time of study) are included in this report. All were born and residing in the United States and were monolingual native speakers of English: they reported learning English and no other languages in the home, and were not fluent in any languages other than English. Participants were recruited via Prolific and received monetary compensation for their time.

Our planned sample size was determined based on a power assessment: We planned to do correlation analyses to test our questions about the reliability of individual variability across phones and tasks using one datapoint per participant, and arrived at a target sample size of 130 by calculating the number of data points needed to detect a relatively weak correlation ($r = 0.3$) with 80% power, using a Bonferroni-corrected alpha-level of $0.05/6 = 0.0083$ (this correction was chosen because we expected to include up to 6 pairwise comparisons in a given analysis). In other words, by collecting data from 130 (or more) participants, we should expect to be able to detect correlations of $r = 0.3$ or greater with 80% power, within our framework for assessing significance.

215 participants initially recruited for the study completed a screening session in which they completed a word-reading task and language background form. 45 of these participants were considered ineligible for the main study because of reported fluency and/or early exposure to other languages ($n = 3$) or recording quality unsuitable for phonetic analysis ($n = 42$). Of the remaining 169 participants who were invited to complete the main study, 18 chose not to participate, and 19 participated but were subsequently excluded because of recording quality, leaving the 133 participants whose data is analyzed here.

### 2.2. Stimuli

The same target stimuli were used in all three tasks (implicit imitation, explicit imitation, and discrimination). Target stimuli consisted of two versions each of 6 "stop words" starting with /p/ or /t/ (*page, poke, pub, tape, tub, type*) and 6 "vowel words" containing /æ/ or /u/ (*map, sad, sack, boot, food, mood*). The two versions of each word were manipulated from a single

baseline token to differ minimally in VOT (for stop words) or F2 (for vowel words). The two versions of each stop words differed by 100 ms VOT, centered around the VOT value of each baseline token.[2] The two versions of vowel words differed in F2 by 1 Bark (for /æ/ words) or 1.75 Bark (for /u/ words), centered around the F2 value of each baseline token.

Our rationale for the specific VOT and F2 differences chosen for this study was based on an effort to equalize perceptual salience across the target contrasts: we wanted the two versions of each word to have similar discriminability, as well as above-chance but below-ceiling performance. We had done previous pilot work in the lab testing discrimination of different magnitudes of difference for each dimension, and chose VOT and F2 differences based on what we expected would elicit the desired perceptual patterns. VOT and F2 values for all stimuli are given in the supplementary materials.

Baseline recordings of all target words were made by a female native speaker of Standard American English and scaled to 70 dB. For stop words, VOT of the natural production was measured, and the two target tokens were created by reducing VOT by 50 ms for the low condition and increasing it by 50 ms for the high condition, using Praat's built-in PSOLA algorithm (Moulines & Charpentier, 1990). For vowel words, F2 of the vowel of each word was manipulated with Praat's built-in LPC synthesis function. F2 was manipulated to match the target value (with low and high levels differing by 1 Bark for /æ/ words and 1.75 bark for /u/ words, centered around the natural production mean) across the full vowel. All other formants were left unmanipulated.

The implicit imitation task also included 34 filler words, which were natural productions of monosyllabic English words by the same speaker. Filler words did not contain any word-initial voiceless stops or /æ/ or /u/ vowels. All filler words are provided in the supplementary materials.

### 2.3. Procedure

The experiment was completed fully online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants were asked to use headphones or earbuds to listen, if possible, and their own recording devices. Based on self-report, listening devices included headphones or earbuds (n = 131) or computer speakers (n = 2), while recording devices included a built-in computer microphone (n = 35), a microphone on earbuds or headphones (n = 75), or an external microphone (n = 23). Participants completed three tasks in the following order: 1) an implicit word imitation task (disguised as a memory study), 2) an explicit word imitation task, and 3) a discrimination task. Both implicit and explicit imitation were elicited via word repetition tasks, a design choice made with two considerations in mind. First, we wanted to keep the two tasks as similar as possible, such that any differences in results could be attributed to the implicit vs. explicit element of the task, as opposed to other task-related differences. At the same time, we wanted to maximize the implicit vs. explicit nature of the tasks, both in terms of the goals (such that participants were clear in the explicit task that the goal was imitation, and thought that the goal was something other than imitation in the implicit task) and in terms of whether attention was explicitly directed to the target phonetic differences. Task order was held constant, with the explicit task and discrimination task always presented after the implicit task, in order to avoid undermining the implicit nature of the latter by drawing attention to the target features and the fact that imitation is a goal of the study.[3]

Fig. 1 shows a schematic of all three tasks. The same set of 24 target stimuli (12 words * 2 levels) were used for all three tasks, and each imitation task presented the full set 4 times, resulting in 96 imitated target tokens per participant per task. The full session took about 30 min to complete. A language background questionnaire had been completed in a prescreening session.

#### 2.3.1. Implicit imitation task

In the implicit imitation task, participants completed four blocks with two phases each: a Listen and Repeat Phase, where participants listened to words and repeated them out loud, followed by a short Memory Test phase, where participants were shown a word and asked whether they had heard it in the study (Fig. 1).

Each of the four test blocks consisted of 48 Listen and Repeat trials, presented auditorily, followed by 4 Memory Test trials, presented visually. The Listen and Repeat trials were designed to test for implicit imitation. On each trial, participants heard a word and repeated it out loud, with the screen advancing automatically after 4 s. The Listen and Repeat trials alternated between target and filler words, with the full set of 24 target stimuli (12 words * 2 levels) included in each block. The Memory Test trials were included to disguise the true purpose of the task. In each Memory Test trial, participants saw a word in English orthography and were asked whether they had heard it during the study. They responded with the keyboard and were given feedback on accuracy. All Memory test words were filler items, 2 of which had been heard and 2 of which had not. Although targets and fillers always alternated, the order of targets and the order of fillers were randomized. An initial practice block contained two Listen and Repeat trials and two Memory Test trials.

#### 2.3.2. Explicit imitation task

In the explicit imitation task, participants listened to pairs of target items differing minimally in the target feature (e.g. 'boot' with low and high F2) and were then asked to imitate the difference (Fig. 1). In each trial, participants listened to a pair of words, then immediately heard each word again but were asked to repeat each word out loud after they heard it, imitating the way it was said. They were specifically asked to pay attention to and try to replicate the differences between the two words.

The task consisted of two blocks that were identical other than the trial-level presentation order: each consisted of 24 trials, or two repetitions of the full set of 12 target pairs. Order of

[2] For stops words, mean VOT was 46 ms and 146 ms for the shortened and lengthened versions, respectively. For vowel words, mean F2 for low and high versions were 1760 Hz (11.98 Bark) and 1992 Hz (12.98 Bark) for /æ/ words and 1309 Hz (10.20 Bark) and 1709 Hz (11.95 Bark) for /u/ words. Original and manipulated values for all tokens are available in the supplementary materials.

[3] While task order is therefore a confound that should be taken into account when interpreting results, we do not have reason to expect that it would affect our questions of interest in a systematic way.

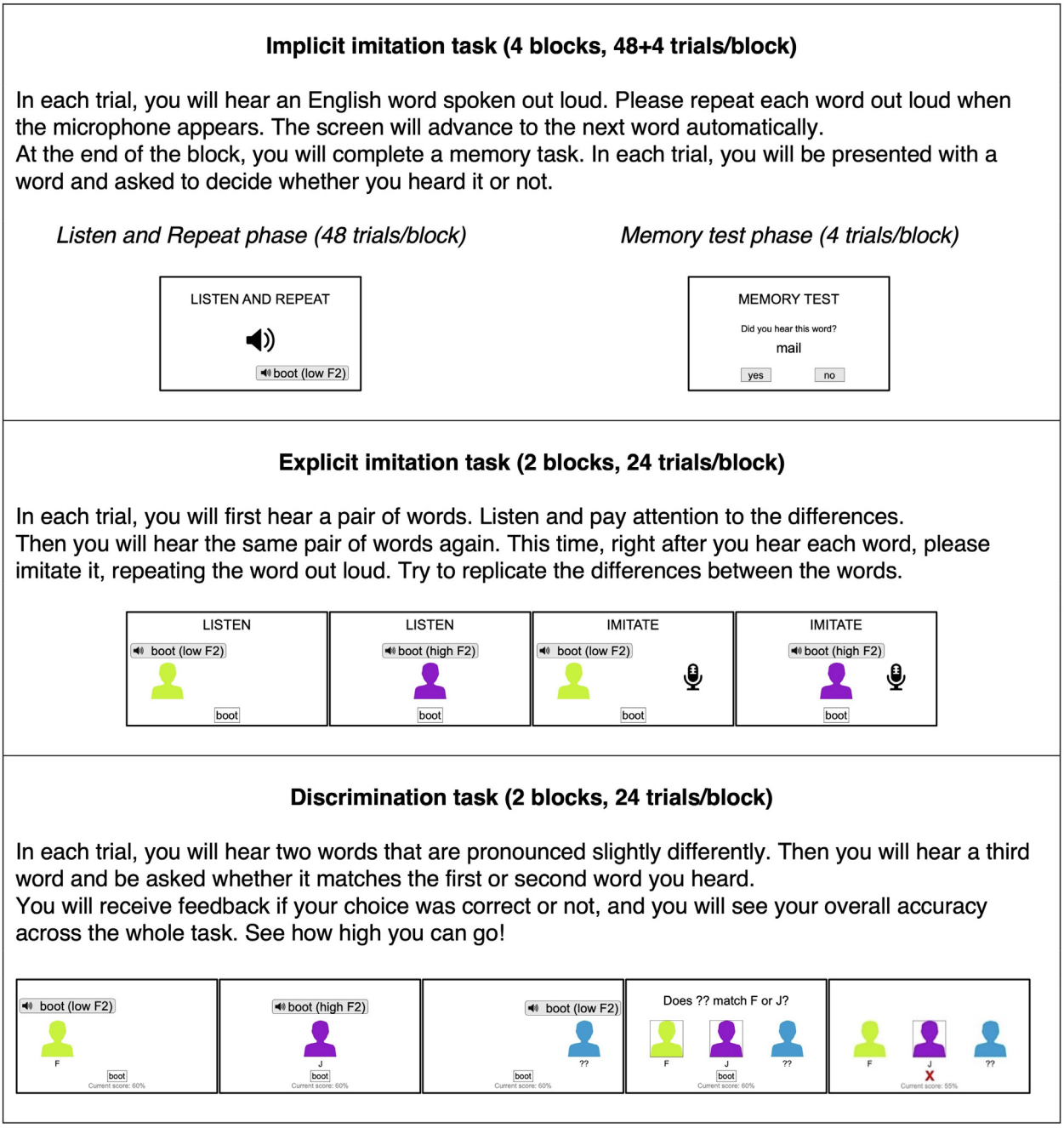


**Fig. 1.** Summary of instructions and procedure for all three tasks. Grey boxes with an audio icon represent the audio clip that was being played to the participant; these were not displayed on the screen.

trials was randomized within each block, and half of the trials had the lower level of the feature presented first (e.g. 'boot' with low then high F2), while the other half had the higher level presented first (e.g. 'boot' with high then low F2). A practice trial preceded the main blocks.

During the listening stage of each trial, the two audio tokens were presented in sequence with a 750 ms break in between. During the Imitation stage, the audio tokens were presented in the same order, but after each one played, a microphone appeared, prompting the participant to imitate the word. The screen advanced automatically after 2.5 s.

### *2.3.3. Discrimination task*

The participants then completed an ABX discrimination task on the same stimuli, in which they were asked to listen to a pair of words, then decide whether a third word they heard matched the first or second word (Fig. 1). The third word was always acoustically identical to the matching word.

The structure was the same as for the explicit imitation task: two blocks of trials, with each block consisting of two repetitions of the 12 target pairs, in randomized order. The three audio clips played on each trial were separated by 750 ms ISI, and a practice trial preceded the main blocks. Feedback was given on each response, as well as the cumulative percentage accuracy across the experiment.

### 2.4. Data processing and acoustic measurements

Each participant had the opportunity to produce 96 imitated target items in each imitation task (12 target words * 2 levels * 4 repetitions), for a total of 192 production per participant. Of the possible 25,536 tokens (192 productions * 133 participants), 401 tokens (1.6% of the data) were omitted because productions were missing (n = 97), contained background or signal noise making annotation impossible (n = 45), contained a word other than the target word (n = 3), or because a reliable formant track could not be identified (n = 256).

All acoustic measurements were done using Praat (Boersma and Weenink, 2021). Production data was segmented at the phone-level using the Montreal Forced Aligner (McAuliffe et al., 2017) to facilitate annotation. VOT of stop words was then annotated automatically with AutoVOT (Keshet et al., 2014), then all tokens were manually checked and corrected if necessary.

For formant measurements for vowel words, each token was manually examined, correcting the force-aligned vowel boundaries if necessary, and selecting parameters that would result in accurate tracking of F2. F2 was then measured at the vowel midpoint. If no accurate track could be identified, the token was excluded from analysis (n = 256). F2 was converted to Bark frequency using a formula from Traunmüller (1990) as implemented in the emuR package (Jochim et al. 2024) prior to analysis.

### 2.5. Statistical analyses

All statistical analyses were performed in R (R Core Team, 2023). Group-level imitation and discrimination were tested with mixed-effects linear (for imitation) and logistic (for discrimination) regression models, using the package lme4 (Bates et al., 2015). P-values for linear models were computed using the lmerTest package (Kuznetsova et al., 2017), and an alpha-level of 0.05 was used as the threshold for significance. In the case of significant interactions, we performed follow-up tests using the phia package (De Rosario-Martinez, 2015). To test the stability of individual variability across features and tasks, we calculated by-participant mean values for the relevant measure and performed correlation tests, using Bonferroni-corrected alpha-levels as a threshold for significance. Details of the model structure and statistical tests vary by task and will be introduced in the relevant subsections of results.

## 3. Results

### 3.1. Group-level effects

In this section, we report group-level results of tests of the presence and extent of convergence in each of the imitation tasks, as well as the perceptibility of the target differences in the discrimination task.

#### 3.1.1. Group-level imitation

Results for both imitation tasks are presented in Fig. 2 and Tables 1 and 2. Fig. 2 shows participants' VOT (for stop words) and F2 (for vowel words) when repeating after target words varying in each dimension, in the implicit and explicit tasks and for each phone separately. Overall, these plots show consistent imitation in both tasks: For each task and for all phones, participants produced higher values of VOT/F2 when repeating after stimuli with high levels of the relevant dimension (light-colored distributions) compared to those with low levels (dark-colored distributions). Furthermore, for each phone, the extent of imitation (as quantified by the difference between the two levels) is always greater in the explicit than in the implicit task.

We used two mixed-effects linear regression models, one for stop words and one for vowel words, to test for the presence of imitation and whether the extent of imitation differed by task and/or phone. Results of these models are shown in Tables 1 and 2. The two models had the same structure and included the predictor variables Task (*Implicit* vs. Explicit), Phone (*/p/* vs. /t/ for the stop model; */æ/* vs. /u/ for the vowel model), Level (*Low* vs. High), and all interactions. Random intercepts were included for Participant and Word, as well as a by-participant random slope for Level.[4] Predictor variables were centered (−0.5, 0.5) and reference levels are in italics above.

We are primarily interested in the magnitude of imitation, or how much participants' productions differ when repeating after stimuli with low vs. high VOT/F2, which corresponds to the effect of Level in our statistical model. Therefore, we structure our interpretation of the models around the effect of Level: the estimate and corresponding p-value for Level gives the extent and significance of imitation overall, across all of the predictor variables, and higher-order interactions between Level and other predictor variables indicate how this effect differs depending on Phone and/or Task.

For stop words, the significant effect of Level indicates that participants showed imitation overall. Significant interactions between Level and Phone and between Level and Task indicate that the extent of imitation is greater for /t/ than /p/ words, and greater in the explicit than in the implicit task. Follow-up tests were done to confirm that the effect of imitation was significant even in the least imitative condition, i.e. implicit imitation of /p/ words ($\chi^2$ = 18.652, p < 0.001). The average amount of imitation was 6 ms for /p/ and 10 ms for /t/ in the implicit task, and 27 ms for /p/ and 28 ms for /t/ in the explicit task.

Results for vowels were qualitatively similar, with a significant overall effect of Level, significant two-way interactions of Level with the other two predictors, and a significant three-way interaction. Follow-up tests again confirmed the significance of imitation in all Phone/Task combinations, including the least imitative, i.e. implicit imitation of /ae/ ($\chi^2$ = 4.31,

[4] This random effects structure was chosen over more complex structures which did not converge; however, the models that we considered theoretically the most motivated, which included by-participant slopes for all predictor variables and their interactions, showed the same qualitative effects.

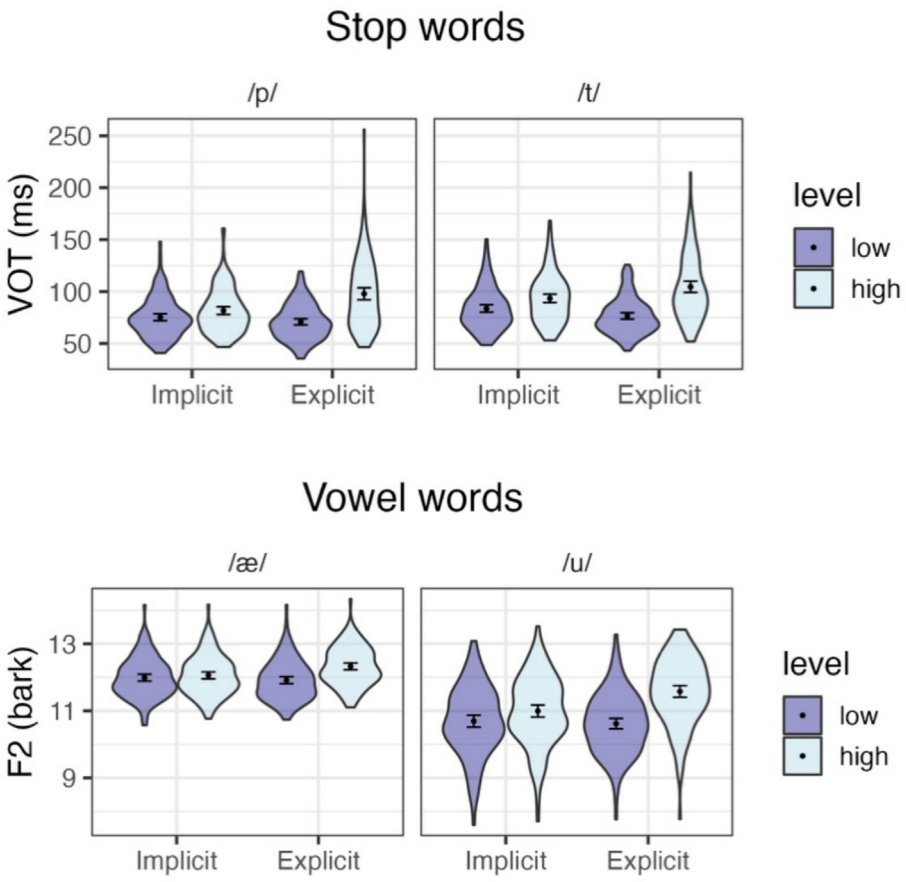


**Fig. 2.** Group-level results for implicit and explicit imitation tasks for each phone separately, showing participants' VOT (for stop words; top) and F2 (for vowel words; bottom) when repeating after low (darker) or high (lighter) values of the target dimension. Violin plots show distribution of by-participant means, with error bars showing 95% confidence intervals.

**Table 1**
Statistical results for group-level imitation of stop words. Significant effects ($p < 0.05$) are in bold. SE = Standard Error. Reference levels are in italics.

| | Estimate | SE | t value | p value |
|---|---|---|---|---|
| **Intercept** | **85.506** | **1.970** | **43.407** | **< 0.001** |
| **Level (high vs. *low*)** | **17.814** | **1.274** | **13.977** | **< 0.001** |
| **Phone (/t/ vs. */p/*)** | **8.170** | **2.256** | **3.621** | **0.022** |
| **Task (explicit vs. *implicit*)** | **3.921** | **0.436** | **9.003** | **< 0.001** |
| **Level (high vs. *low*) * Phone (/t/ vs. */p/*)** | **2.215** | **0.871** | **2.543** | **0.011** |
| **Level (high vs. *low*) * Task (exp. vs. *imp.*)** | **19.472** | **0.871** | **22.355** | **< 0.001** |
| **Phone (/t/ vs. */p/*) * Task (exp. vs. *imp.*)** | **−3.963** | **0.871** | **−4.550** | **< 0.001** |
| Level (high vs. *low*) * Phone (/t/ vs. */p/*) * Task (exp. vs. *imp.*) | −2.304 | 1.742 | −1.323 | 0.186 |

**Table 2**
Statistical results for group-level imitation of vowel words. Significant effects ($p < 0.05$) are in bold. SE = Standard Error. Reference levels are in italics.

| | Estimate | SE | t value | p value |
|---|---|---|---|---|
| **Intercept** | **11.522** | **0.078** | **147.599** | **< 0.001** |
| **Level (high vs. *low*)** | **0.434** | **0.026** | **16.955** | **< 0.001** |
| **Phone (/u/ vs. */æ/*)** | **−1.102** | **0.114** | **−9.690** | **< 0.001** |
| **Task (explicit vs. *implicit*)** | **0.173** | **0.012** | **13.927** | **< 0.001** |
| **Level (high vs. *low*) * Phone (/u/ vs. */æ/*)** | **0.395** | **0.025** | **15.863** | **< 0.001** |
| **Level (high vs. *low*) * Task (exp. vs. *imp.*)** | **0.497** | **0.025** | **19.957** | **< 0.001** |
| **Phone (/u vs. */æ/*) * Task (exp. vs. *imp.*)** | **0.158** | **0.025** | **6.339** | **< 0.001** |
| **Level (high vs. *low*) * Phone (/u/ vs. */æ/*) * Task (exp. vs. *imp.*)** | **0.325** | **0.050** | **6.524** | **< 0.001** |

$p = 0.038$). The average amount of imitation was 0.07 Bark for /ae/ and 0.30 Bark for /u/ in the implicit task, and 0.40 Bark for /æ/ and 0.96 Bark for /u/ in the explicit task.

In sum, the group results for the imitation tasks showed robust imitation effects: participants imitated VOT and F2 in the relevant phones in both implicit and explicit imitation tasks, although, as expected, imitation was consistently greater in the presence of explicit instructions to imitate. There were also overall phone-related differences in imitation, with more imitation of /t/ than /p/, and more imitation of /u/ than /ae/. However,

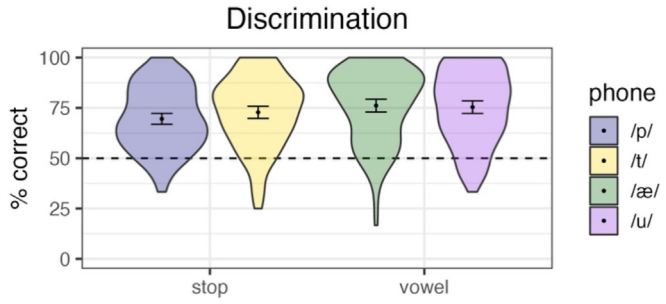


**Fig. 3.** Group-level results for the discrimination task. Violin plots show distribution of by-participant accuracy in discrimination of each phone, with error bars showing 95% confidence intervals.

given the relatively small number of words used for this task, we do not want to overgeneralize about these phone-related differences.

*3.1.2. Group-level discrimination*

A discrimination task was included primarily as a method of testing the relationship between an individual's extent of imitation and their ability to perceive the target difference. Group-level results for this discrimination task are in Fig. 3 and Table 3. Fig. 3 shows the distribution of by-participant mean accuracy in discrimination of target words differing in VOT (for stop words) or F2 (for vowel words), shown separately for each phone. Overall, these plots show robust above-chance performance on a group level, with a wide range of variability in individual performance.

We tested whether discrimination accuracy differed by phone using a mixed-effects logistic regression model. The binary outcome variable was Accuracy on each trial, and the fixed predictor was Phone (/p/, /t/, /æ/, /u/, dummy-coded with /p/ as the reference level). Random intercepts were included for Participant and Word (a model including a by-participant random slope for Phone was did not converge but showed the same qualitative results).

Statistical results are shown in Table 3. The significant intercept indicates above-chance performance for /p/ words (i.e. at the reference level), and the significant effects of Phone indicate that accuracy was higher for /æ/ and /u/ than it was for /p/. A model testing the difference between adjacent levels (using backward-difference coding of Phone) showed that there was no significant difference between accuracy between /t/ and /æ/, or between /æ/ and /u/. Mean accuracies were 70% for /p/, 73% for /t/, 76% for /æ/, and 75% for /u/.

In sum, performance was well above chance (70–76% accuracy) for all featural differences tested in this study. Accuracy on the two vowels was significantly higher than /p/ words (while /t/ words were intermediate but did not differ significantly from either /p/ or vowel words), but overall, phone-based differences were fairly small in magnitude.

*3.2. Individual variability*

We structure the analysis of individual patterns in terms of the three questions laid out in the introduction: 1) How stable are individual patterns of imitation across phones with different degrees of phonetic similarity?; 2) How well is individual variability in imitation predicted by discrimination ability?; and 3) How stable are individual patterns across implicit and explicit imitation? We perform these analyses using indices of each participant's imitation or discrimination for the relevant phone (s)/task, then perform Pearson's correlation tests, using a Bonferroni-corrected alpha-level to assess significance.

Before presenting the correlation analyses, we first present a summary of individual patterns across both imitation tasks. For each task and each phone, we calculated an individual "imitation index" for each participant: the difference in VOT duration (for stop words) or F2 (for vowel words) when comparing their repetition of "high" vs. "low" levels, such that a positive value represents imitation, with higher values representing more imitation.[5]

Fig. 4 shows the distribution of individual imitation indices for each phone separately. Overall, these reiterate the patterns discussed above in the group results: there was imitation for all of the target phones in both tasks, with most individuals showing imitation (i.e. positive values) in all phones for both implicit and explicit imitation tasks, with more imitation in the explicit than in the implicit task, as indicated by higher imitation indices in the explicit task for all phones). However, this graph reveals one important property of the results that was not discussed above: not only is there more imitation in the explicit task, but there is also a much greater range of individual variability: aggregated across both phones of a given class, the standard deviation of the stop imitation indices was 3.87 times greater in explicit (SD = 27.10) than implicit (SD = 7.01) tasks, and for vowels, variability was 3.23 times greater in explicit (SD = 0.533) than implicit (SD = 0.165).

*3.2.1. Patterning of individual variability in imitation across phones*

We first test whether patterns of individual variability in imitation generalize across different phones, and whether correspondences are stronger across sounds sharing in the same class (e.g. a stronger correlation between /p/ and /t/ than between /p/ and /æ/). Since implicit and explicit imitation tasks involve different processes that might be expected to vary across individuals, we examine this question separately for implicit and explicit imitation tasks.

Results are shown in Fig. 5, which includes scatterplots showing the individual extent of imitation, as well as the results of a correlation test, for each of the six possible pairwise comparisons of the four phones, for implicit and explicit imitation separately, using a Bonferroni-corrected alpha-level of 0.05/6 = 0.008. In the plots, each point represents one participant and shows that participant's amount of imitation for the two phones shown on the x- and y-axes.

In implicit imitation (top row), individual performance was significantly correlated for phones within a class: individual participants' extent of implicit imitation of /p/ predicted imitation of /t/, and individual imitation of /æ/ predicted imitation of /u/. None of the other correlations (i.e. those between phones of different classes) were significant.

For the explicit imitation task, all six pairwise correlations were significant. The correlations were strongest (as evi-

[5] In some analyses, as will be discussed below, the imitation index is calculated over a larger domain; e.g., a combined index of both stops together. In these cases, the calculation is done in the same way: the difference between a participants' mean value when repeating the higher vs. the lower level of the target stimuli.

**Table 3**
Statistical results for the discrimination task. Significant effects (p < 0.05) are in bold. Reference levels are in italics.

| | Estimate | Standard error | z value | p value |
|---|---|---|---|---|
| **Intercept** | **0.897** | **0.112** | **7.974** | **< 0.001** |
| Phone (/t/ vs. /p/) | 0.170 | 0.141 | 1.203 | 0.229 |
| **Phone (/æ/ vs. /p/)** | **0.361** | **0.142** | **2.539** | **0.011** |
| **Phone (/u/ vs. /p/)** | **0.316** | **0.142** | **2.224** | **0.026** |

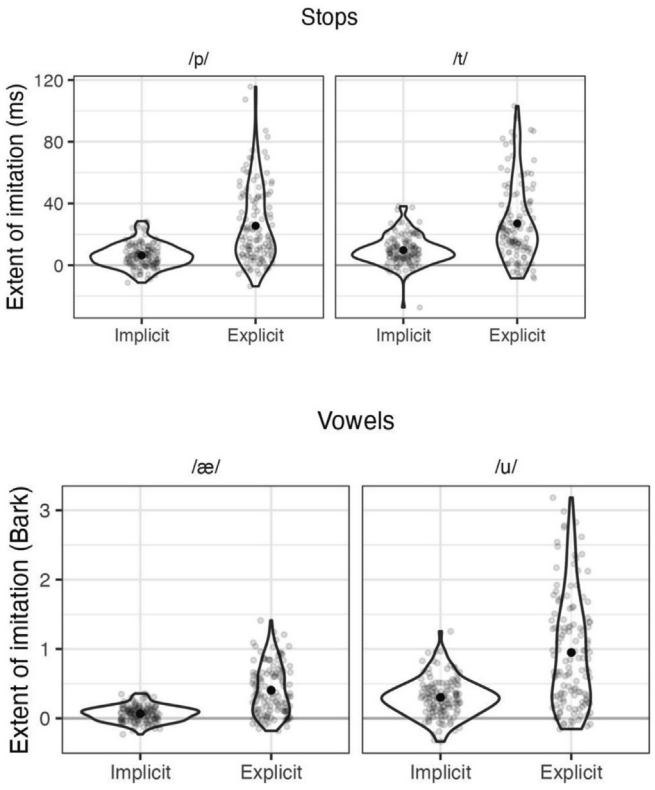


**Fig. 4.** Distribution of individual extent of imitation for each phone in implicit and explicit imitation. Each point represents an individual participant's "imitation index" for each feature, i.e. the difference in VOT (for stops) or F2 (for vowels) when repeating after high vs. low levels of the target feature. Black dots show the group mean for each distribution.

denced by the highest r values) for the within-class comparisons: $r$ = 0.855 and $r$ = 0.624 for the stop comparison and vowel comparison, respectively, as compared to $r$s ranging from 0.415 to 0.480 for the different-class comparisons. Notably, the correlation was stronger for all of the between-phone comparisons in the explicit task than any of those from the implicit task: the lowest correlation for explicit imitation ($r$ = 0.415 for /t/ vs. /æ/) was still higher than the strongest (within-class) correlations in implicit imitation, which both had $r$s of 0.344.[6]

In sum, there were clear correlations between extent of individual imitation of related phones in both the implicit and explicit imitation tasks. The extent of individual imitation for unrelated phones was also correlated in an explicit imitation task, although the correspondence was weaker than it was for related phones in the same task. On the other hand, the correlations for unrelated phones were not significant for implicit imitation, though all correlations were numerically positive.

[6] The same correlation analyses were run on a dataset excluding one participant who had very high explicit imitation scores (>150ms for both /p/ and /t/). Removal of this participant did not change the direction or significance of effects, and did not affect r-values by more than 0.05 for any comparison.

*3.2.2. Patterning of individual variability across implicit vs. explicit imitation tasks*

We now turn to the comparison of individual performance across the two imitation tasks, examining to what extent an individual's explicit imitation ability is related to the extent to which they show imitation in an implicit task. We test the correspondence between the two types of tasks for vowels and consonants separately, using individual participant imitation indices, calculated in the same way as above.

Scatterplots of individual extent of imitation on the two tasks, and results of the correlation tests, are shown in Fig. 6. First, these graphs again highlight the difference in extent of individual variability across the two tasks: while imitation was, across the board, greater in the explicit task, as shown by the higher numbers overall, another important difference was that participants showed a larger range of extent of imitation in the explicit than the implicit task. In terms of the correlation between tasks, individual imitation scores were weakly but significantly correlated across implicit and explicit imitation for both stops and vowels (r = 0.217 and 0.188 respectively). In sum, there appears to be a relationship between individuals' convergence on the implicit imitation task and how much they imitated on the explicit task, but this relationship is quite weak.

*3.2.3. Role of discrimination ability in predicting imitation*

Finally, we examine the question of whether individual differences in imitation might be attributable to differences in perception, testing the correspondence between each individual's imitation and their discrimination accuracy. We test this correspondence separately for vowels and consonants, using individual participant imitation indices calculated as above (i.e. the difference between imitation of "high" and "low" levels of the relevant stimuli), but combining the two phones in each class (i.e., each participant has a single value for explicit imitation of stops), and using each individual's mean accuracy for discrimination of each class separately as a discrimination index.

Fig. 7 shows scatterplots of the relationship between individual imitation and discrimination accuracy, for stops and vowels separately. The patterns are qualitatively similar for both stops and vowels: Explicit, but not implicit, imitation performance is significantly correlated with discrimination accuracy, and these correspondences are fairly strong (r = 0.590 and 0.621 for stops and vowel respectively). In sum, there is a clear relationship between explicit imitation and perception in the tasks tested here, but no evidence that individual differences in implicit imitation are attributable to different perceptual strategies.

We also tested whether stop and vowel discrimination were correlated on an individual level. We found a significant correlation (r = 0.410, p < 0.001), indicating that individuals who were more accurate at discriminating the stop contrasts were also more accurate at vowel discrimination, as shown in Fig. 8.

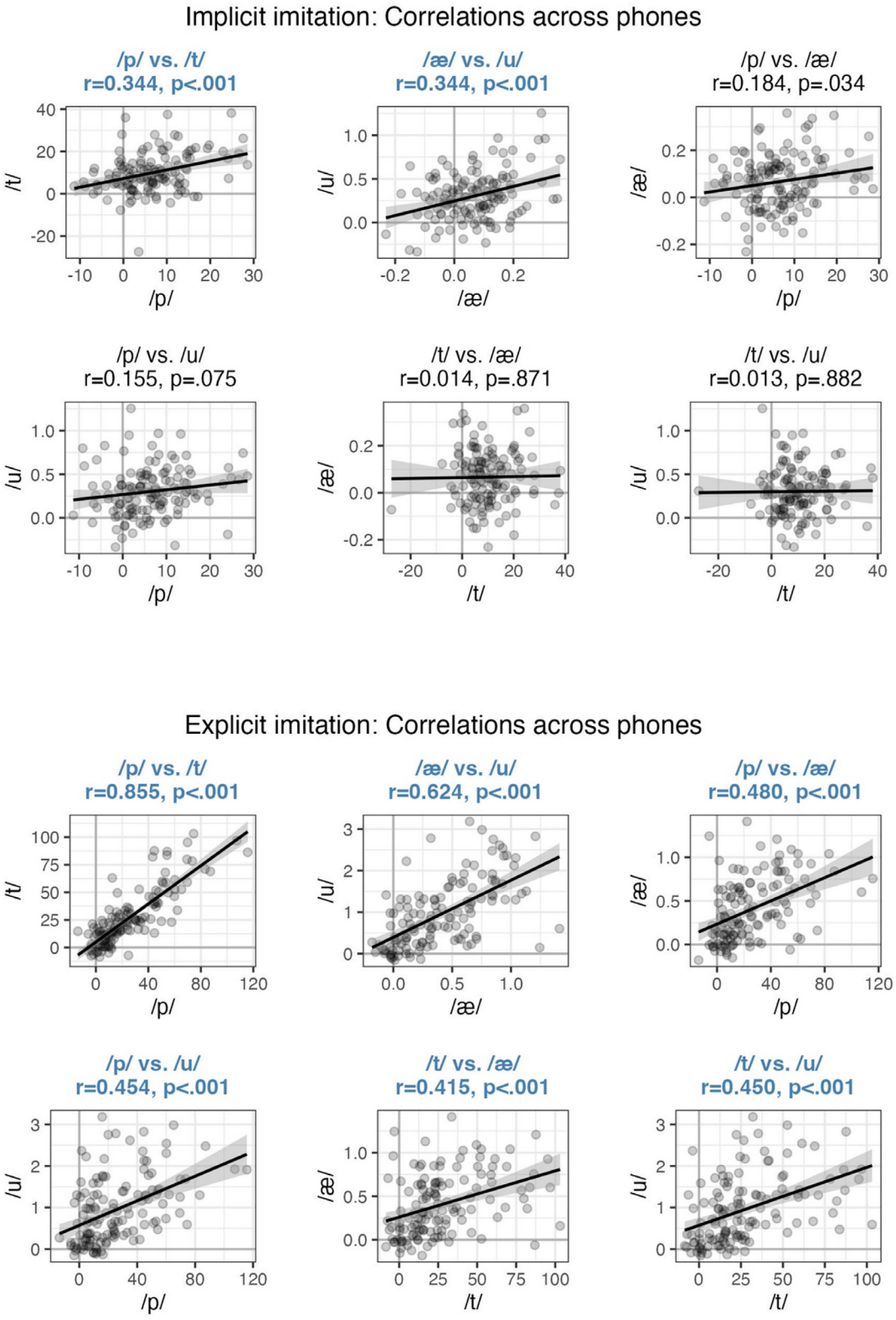


**Fig. 5.** Scatterplots showing individual extents of imitation for all pairwise comparisons of four phones in implicit (top) and explicit (bottom) imitation tasks. Each plot shows each individual's imitation index for the two relevant phones, the best-fit regression line and 95% confidence interval, and the results of a Pearson's correlation test. Significant results are shown in bold and colored font.

The fact that discrimination of stops and vowels is related brings up the possibility that the finding of the across-class correlation between explicit imitation of stops and vowels, discussed above, might be due not to any meaningful relationship between *imitation* across the classes, but rather simply to the fact that *discrimination* of the two classes is correlated, and imitation follows from discrimination. We therefore tested whether the significant individual-level correspondence

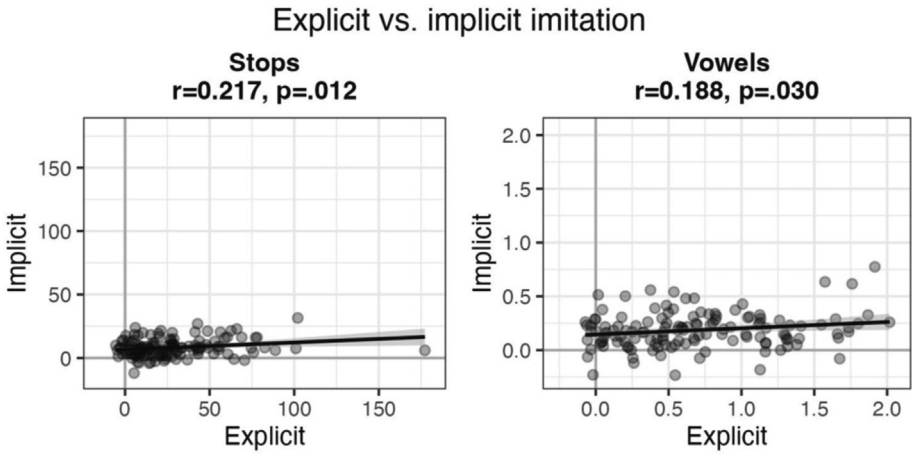


**Fig. 6.** Scatterplots showing the relationship between individual performance on the implicit vs. explicit imitation tasks, for stops (left) and vowels (right). Each point represents an individual, and the plot shows best-fit regression lines, 95% confidence intervals for the slopes, and the results of a Pearson's correlation test of each relationship.

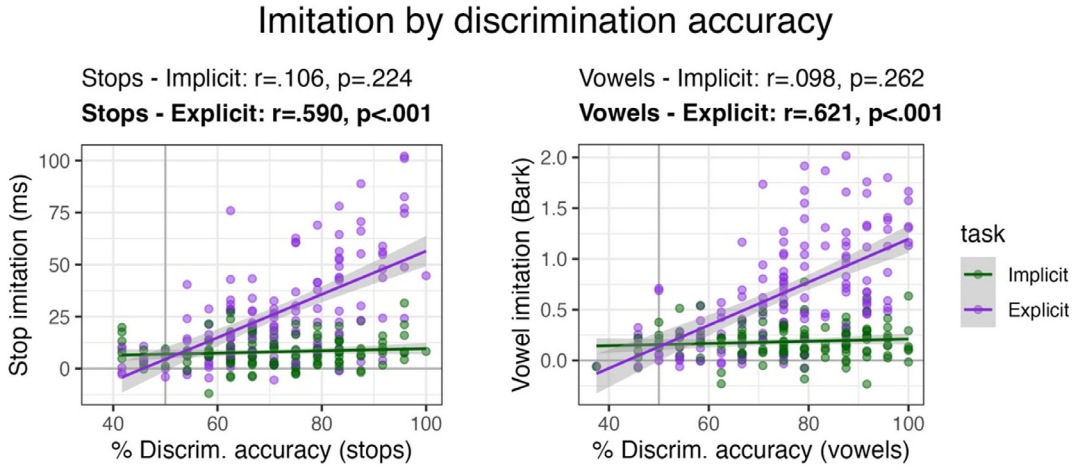


**Fig. 7.** Scatterplots showing the relationship between individual discrimination and imitation, for stops (left) and vowels (right). Each point represents an individual, representing their discrimination accuracy on the x-axis and their implicit (purple) and explicit (purple) imitation index on the y-axis. Best-fit regression lines and 95% confidence intervals are shown, as well as the results of a Pearson's correlation test of each relationship. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

across classes in explicit imitation (i.e. whether explicit stop imitation is related to explicit vowel imitation) holds even when controlling for discrimination, via model comparison. For each class, we built two regression models, the first predicting explicit imitation from discrimination of the same class, and the second adding a predictor of imitation of the different class (Table 4). In both cases, the model containing across-class imitation as a predictor showed significantly improved fit as compared to the discrimination-only model, indicating that explicit imitation of the other class contributed to explicit imitation accuracy above and beyond the role of discrimination.

## 4. Discussion and conclusions

### *4.1. Summary of results*

This work examined phonetic imitation of subphonemic differences in VOT of voiceless stops /p/ and /t/ and F2 of the vowels /æ/ and /u/ in two types of tasks: a word-repetition task (disguised as a memory task) designed to elicit implicit imitation, and an explicit imitation task in which participants heard minimal pairs differing in the target feature and were asked to pay attention to, and imitate, the differences. We tested the stability of individual performance across different phones, features, and tasks, as well as the extent to which discrimination accuracy predicted the extent of an individual's imitation.

On a group level, we found that participants showed imitation of both VOT and F2 differences in both tasks, and that they showed more imitation in the explicit than the implicit task. On an individual level, we found evidence for structured variability in imitation patterns: individual levels of imitation were related in systematic ways across the different target phones and tasks.

Fig. 9 presents a visual summary of the within- and across-task individual correlations tested above. The purpose is to be able to compare the relative strength of each correlation in predicting individual performance for each type of imitation, for stops and vowels separately. The left panel shows how tightly individual variability in explicit imitation is predicted by individ-

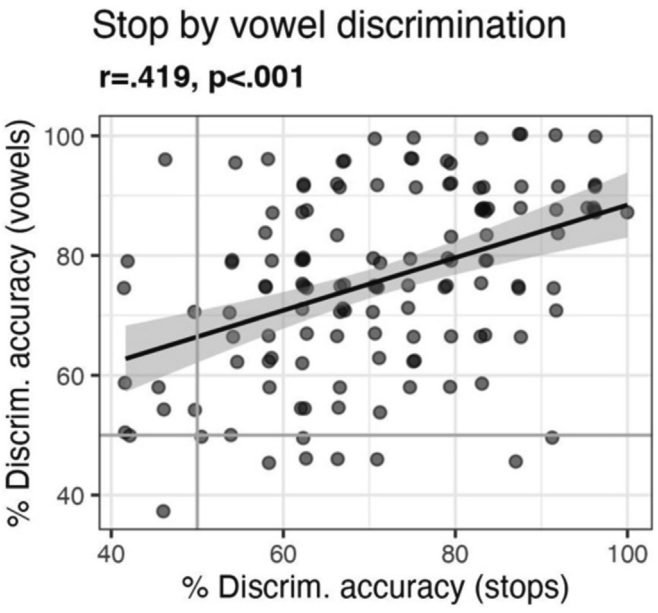


**Fig. 8.** Scatterplot showing the relationship between individual discrimination accuracy for stops vs. vowels. Each point represents an individual, representing their discrimination accuracy for stops on the x-axis and vowels on the y-axis. The best-fit regression line and 95% confidence interval is shown, as well as the results of a Pearson's correlation test of the relationship.

**Table 4**
Statistical models used to test whether cross-class individual patterns of explicit imitation are predictive above and beyond what might be accounted for by discrimination ability.

| | |
|---|---|
| **Stop** | Model 1: lm(Explicit stop imitation ~ Stop discrimination)<br>Model 2: lm(Explicit stop imitation ~ Stop discrimination + Exp. vowel imitation)<br>ANOVA results of model comparison: F = 27.927, p < 0.001 |
| **Vowel** | Model 1: lm(Explicit vowel imitation ~ Vowel discrimination)<br>Model 2: lm(Explicit vowel imitation ~ Vowel discrimination + Exp. stop imitation)<br>ANOVA results of model comparison: F = 19.278, p < 0.001 |

ual performance in 1) explicit imitation of the same class of sounds; 2) explicit imitation of a different class of sounds; 3) implicit imitation; and 4) discrimination. The analogous comparisons for implicit imitation are shown in the right panel.[7]

The first individual-level analysis tested whether an individual's extent of imitation was stable across phones, and whether this stability was greater for phones with the same target feature. We found that in both types of imitation, individuals' extent of imitation did correspond across phones within the same class (individuals who imitated VOT differences in /p/ more than average also imitated /t/ more than average, and those who imitated F2 differences in /æ/ more than average also imitated /u/ more than average), as shown by the above-zero correlations corresponding to the "within-class" bars in both panels of Fig. 9. When considering imitation of phones from different classes (i.e. stops vs. vowels), we also found a significant, albeit weaker, correspondence in the explicit imitation task: individuals who imitated VOT differences more than average also imitated F2 differences more than average. In implicit imitation, on the other hand, there was no evidence of correspondence between individuals' imitation of stops and vowels. In other words, while we saw evidence of structured variability in both imitation tasks, the extent and nature of individual reliability depended on the task: the correspondences were both less reliable (as shown by weaker correlations overall) and less generalizable (as shown by the lack of significant correlation in the across-phone analyses) in implicit, as compared to explicit, imitation.

We also examined the relationship between implicit and explicit imitation performance. We found a significant correspondence between individual performance in the two types of imitation (shown twice in Fig. 9, once in each panel): individuals who showed a high degree of explicit imitation also showed greater-than-average spontaneous imitation, though this correspondence was quite weak.

Finally, we tested whether individual patterns of imitation depended on differences in discrimination ability. Again, we found that the answer to this differed for the two imitation tasks: an individual's discrimination accuracy was a fairly strong predictor of individual imitation in the explicit task, but there was no detectable relationship between perception and imitation in the implicit task.

Along with these primary questions, the summary in Fig. 9 allows us to compare the magnitude of the correspondences for each type of imitation. In both implicit and explicit imitation, individuals' imitation of phones within the same class was the strongest predictor for the extent of individual imitation. For explicit imitation, discrimination accuracy was the next strongest correspondence, followed closely by across-class imitation, with implicit imitation indices showing the weakest correspondence. For implicit imitation, this same correspondence (between explicit and implicit imitation) was the second most predictive, with discrimination and across-class imitation failing to predict implicit imitation patterns. The relative ranking of the different correspondences was consistent across stops and vowels.

### *4.2. Uniformity of phonetic imitation across features and tasks*

The broad question motivating this work was the generalizability of individual differences in imitation: in other words, are some individuals globally greater imitators than others, or is imitation feature-specific – and are individual patterns consistent across implicit and explicit tasks? The few existing studies directly examining the reliability and uniformity of individual variability in imitation have resulted in mixed findings. On the one hand, Wade et al. (2021) found consistent individual performance on imitation of a single feature (extended VOT) across multiple sessions of an identical task, providing important confirmation of within-task individual reliability. However, studies testing consistency across different tasks or features have found no correspondence (different tasks: Sato et al., 2013; different features: Sanker, 2015; Weise & Levitan, 2018; Cohen Priva & Sanker, 2020). The current study was an attempt to provide a controlled test of cross-feature uniformity in two different types of imitation tasks.

[7] These correlations were calculated in the same way as in the analyses above, based on individual indices of imitation (or discrimination) over the relevant subset of data, though sometimes in a more aggregated form than the above for clarity of visualization so as to give a single value for each relevant predictor.

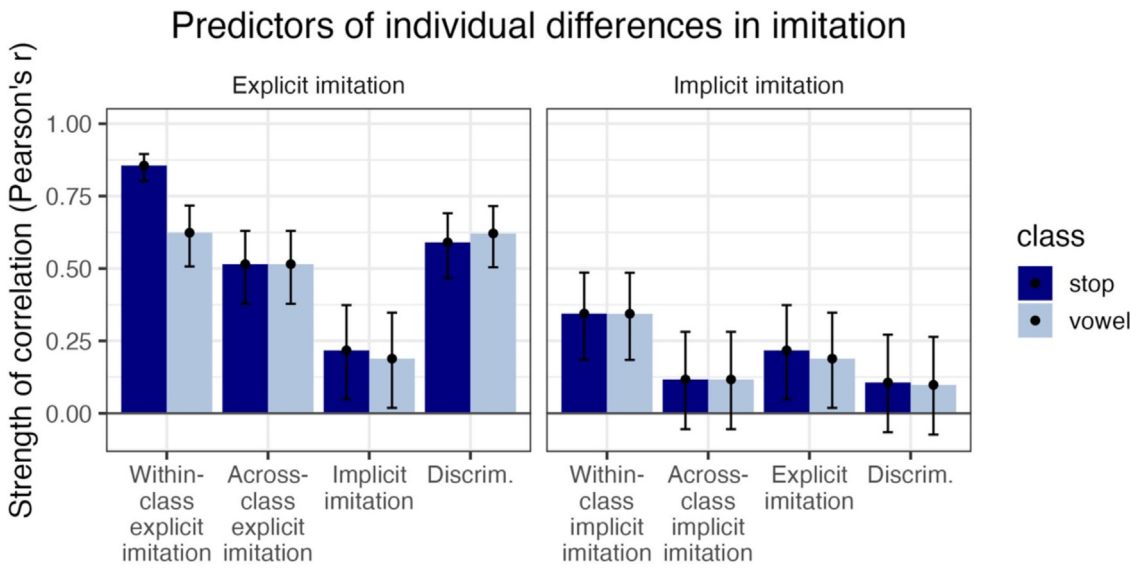


**Fig. 9.** Summary of predictors of individual patterns of explicit (left panel) and implicit (right panel) imitation. Each bar shows the strength of the correlation (Pearson's r) and 95% confidence interval of the estimate, based on the correlation between the two individual indices of performance.

We found that the extent of uniformity across features (VOT and F2) differed for the two imitation tasks. In the explicit task, individual performance was related across features, consistent with the idea that some individuals are indeed "better (explicit) imitators," regardless of the specific feature being imitated. On the other hand, no cross-feature consistency was found in the implicit imitation task. These results are in line with those from previous work finding no clear correspondence between individual extents of imitation of different features in spontaneous imitation in conversational speech (Sanker, 2015; Weise and Levitan, 2018; Cohen Priva & Sanker, 2020), and add to this body of work by showing that this lack of correspondence holds even when using a highly controlled task. However, despite the lack of cross-feature consistency in implicit imitation, individual performance was still characterized by structured variability on a finer-grained level. We found that an individual's extent of imitation was related for phones *within* a class in both types of imitation: the amount that an individual imitated VOT (relative to other participants) was related across different stops, and the amount of F2 imitation was related across vowels.

The differential patterns of cross-feature uniformity across implicit and explicit tasks indicate that the factors governing individual imitation performance are different for the two types of imitation; this is further supported by the relatively large amount of non-shared variance in correlations of individual indices of imitation across the two tasks. Here we consider what our results suggest about the types of individual predictors that may be relevant for each type of imitation.

First, it appears that some subset of non-phonetically-selective factors play a meaningful role in predicting individual variability in explicit imitation. We think it likely that this is due to task-related factors such as motivation, or willingness or ability to intentionally modify production norms in a substantive way. However, general cognitive or personality traits, attitudes toward the talker, or perception- or production-based factors applying across the board (i.e., any factor expected to affect all sounds equally) could equally be driving this uniformity.

At the same time, whatever these non-phonetically-selective factors underlying the uniformity in explicit imitation may be, they do not appear to predict individual variability in implicit imitation. There are different explanations for why this could be the case. It is possible that a factor important for explicit imitation is simply not relevant for implicit imitation: for example, as discussed in the Introduction, individual differences in task motivation are more likely to condition variability in an explicit task where imitation is the stated goal, than in one where it is not. On the other hand, it is possible that a factor could be equally relevant for both types of imitation, but that there are other, phonetically-selective, factors that play a more important role in implicit imitation, outweighing its effect. For example, in the introduction we posited that individual differences in dimension-specific cue weighting or attention might play a greater role in implicit than explicit imitation. If so, then we would expect to find a larger cross-feature disconnect in individual performance in implicit than in explicit imitation, even if there are other general factors playing an equal role in both.

Despite the fact that there is a clear disconnect in performance across the two tasks, it is important to keep in mind that there was a significant, albeit small, relationship between the two: those individuals who showed more spontaneous convergence showed, on average, more imitation in the explicit task as well. This is the first study, as far as we are aware, to find a detectable relationship between implicit and explicit imitation, providing empirical support for the idea that the two types of imitation draw on shared processes (as proposed by Dufour & Nguyen, 2013; Schertz et al., 2023).

In sum, the current work provides empirical support for models of implicit and explicit imitation that rely on both shared and distinct processes. Our specific patterns of results further indicate that there is a relatively large disconnect between individual performance on implicit and explicit imitation tasks, and that this disconnect is likely driven by the fact that that non-phonetically-selective factors play a proportionally larger role in explicit than in implicit imitation. Determining the provenance of both the shared components and the components where the two types

of imitation differ can be addressed in future work, using targeted experiments that examine the independent role of specific predictors in implicit and explicit imitation. This line of work will be useful in laying the groundwork for accurate models of the cognitive processes governing each type of imitation.

### 4.3. Provenance of within-class consistency

The within-class consistency (i.e., the fact that individual imitation indices for /p/ and /t/ are related, as are those for /æ/ and /u/), found in both explicit and implicit imitation, is consistent with the idea that certain individuals may be more prone to imitate certain features – both spontaneously and in explicit imitation tasks. This is related to the idea of target uniformity, or the tendency for individual production patterns to be consistent across related phones; for example, individuals who have longer-than-average VOT for /p/ also tend to have longer-than-average VOT for /t/ (Chodroff & Wilson, 2017, see also Chodroff & Wilson, 2022; Theodore et al., 2009). Our results extend findings from this previous work to show that feature-specific individual tendencies apply not only to baseline production patterns but also to imitation.

It is worth noting that this within-class consistency was present in two features/classes that are qualitatively quite different: VOT of voiceless stops /p/ and /t/, and F2 of the vowels /æ/ and /u/. Voiceless stops are uncontroversially considered a phonological and phonetic class in English, and the primary defining feature, VOT, has been shown to pattern consistently across places of articulation within individuals in production studies (e.g. Chodroff & Wilson, 2017). Given this, it is perhaps unsurprising to find that individuals' imitation patterns are also consistent within this class of sounds. On the other hand, the status of the vowels /æ/ and /u/ as a unified class is a bit more tenuous, and they also are characterized by substantial dialectal and allophonic variation (Jacewicz et al., 2011). Given this, the finding of consistency in imitation of F2 for these two phones is particularly striking.

What constitutes a "class," or the domain of consistency, is an open question. While here we have targeted two specific acoustic features, VOT and F2, it is likely that the domains of consistency are broader. For example, some listeners might show more imitation of spectral characteristics, while others are more imitative of temporal characteristics, or listeners could differ in their imitation of segmental vs. suprasegmental parameters; both of these are distinctions that have been proposed to underlie individual differences in speech perception more generally (e.g., Makashay, 2003; Symons et al., 2023). Existing work on imitation does not provide much evidence bearing on the question of what constitutes the domain of consistency. Studies have included measures that might be expected to pattern together (e.g. imitation of f0 mean and f0 range in Cohen Priva & Sanker, 2020; F1 and F2 in Sanker, 2015) have found no correspondences, even between these closely related measures; however, these studies were examining spontaneous convergence in conversational interaction, where the extensive variability inherent to the task could make the relationship difficult to detect. Therefore, it may be useful to carry out more controlled studies across a range of measures or features in order to determine the scope of consistency of individual imitation performance.

This within-class consistency could, in principle, arise from production and/or perception-based mechanisms. From a production point of view, speakers could differ in how articulatorily malleable or variable a specific dimension is, with greater flexibility corresponding to more imitation. In terms of perception, it seems intuitively plausible that individual differences in the extent of reliance on, or attention to, specific phonetic dimensions, might correspond to their extent of imitation of those dimensions. However, as far as we are aware, there is not existing evidence supporting either of these. While several studies have found a correspondence between articulatory flexibility or ability and imitation, these have all been on a general level (Reiterer et al., 2013; Lee et al., 2021; Myers et al., 2024). For perception, the one study examining this directly (Kim & Clayards 2019) did not find a straightforward relationship between feature-specific perception and imitation, and results from the current work also do not show a relationship between discrimination and implicit imitation, as will be discussed in the following section. Determining what constitutes a "class" as well as the provenance of this within-class consistency, and whether it is driven by perception, production, or a combination, are therefore open questions to explore in future work.

### 4.4. Role of perception

Perception is a fundamental component of imitation: if a phonetic difference is not auditorily perceptible (i.e. below the threshold of a just noticeable difference), imitation is not possible. However, it is also likely that perceptual factors beyond this basic ability to auditorily perceive the difference play a role in imitation, whether it be the amount of attention directed to a given dimension, the phonological importance of a dimension, or different levels of sensitivity to a given dimension (see Schertz et al. 2023 for discussion). Examining the precise role of perception in imitation was not the focus of the current work, but we did want to ensure that the differences to be imitated were indeed perceptible, and therefore tested participants' discrimination of the acoustic variation present in the stimuli used for the imitation tasks. These tests confirmed that overall, all differences were perceptible on a group level, and we further took the opportunity to examine whether individual variability in discrimination accuracy predicted performance on the imitation tasks. We found that the strength of discrimination as a predictor depended on the type of imitation. Discrimination performance was a fairly strong predictor of individual variability in the explicit imitation task, but was not predictive of implicit imitation: instead, the range of individuals' implicit imitation was remarkably consistent across the full range of discrimination accuracy.

The relationship between discrimination and explicit imitation, i.e., the fact that individuals who were less accurate in identifying phonetic differences were also less prone to imitate them, is fairly unsurprising, particularly given that the presentation context for these two tasks was similar: hearing a minimal pair of sounds in direct sequence. It is also possible that external factors, like task engagement, contributed to this relationship: participants less likely to make an effort on a discrimination task might also be less likely to make an effort in accurate imitation. Trying to tease apart the roles of these

factors, and pinpointing the source(s) and directionality of the relationship between discrimination and imitation, could inform L2 sound training paradigms (Nagle & Baese-Berk, 2022; Henderson & Rojczyk, 2023).

While no significant relationship was found between discrimination and implicit imitation, this does not necessarily mean that perception plays less of a role in spontaneous than in explicit imitation. Instead, we think it likely that perceptual factors play a role in both types of imitation, but that these perceptual mechanisms may be different. In our design, the procedure of the discrimination task was closely linked to the explicit imitation task. In both discrimination and explicit imitation, participants were prompted to pay attention to and identify a minimal difference between two sounds presented back to back, and were provided feedback on their choice in the discrimination task. On the other hand, in the implicit task, target words were not presented with their minimal pair, but rather in isolation, in the context of other filler words. As discussed above, implicit imitation requires tracking (and reproducing) a specific value of a cue when heard in isolation, something that was not tested by our discrimination task. It is therefore possible that a different sort of perception task which tests the importance of a cue in listeners' representations, or their sensitivity to fine-grained differences on a given dimension, might be more likely to correspond to implicit imitation.

### 4.5. Methodological implications

Knowing the extent of reliability of individual metrics of imitation on a given task (e.g. Wade et al., 2021) and to what extent an individual's imitation in one feature or task can be taken as a proxy for imitation more generally, has practical implications in terms of research design and interpretation.

The first issue involves the reliability of individual acoustic metrics used as a proxy for imitation more generally. Our results support and build on findings reported in previous work (Sanker, 2015; Weise & Levitan, 2018; Cohen Priva & Sanker, 2020) that found that individual patterns of spontaneous imitation were not stable across different phonetic features. This previous work examined implicit imitation in conversational speech, which is highly variable both in terms of the linguistic content and the phonetic properties of the exposure speech. Because of this, small convergence effects may be unlikely to be found. The current work showed that this disconnect between imitation of two phonetic features (VOT and F2) held even in a highly controlled paradigm. This feature-specificity in implicit imitation underscores the fact that imitation of a single feature should not be taken as a proxy for an individual's overall tendency to imitate, underscoring previous claims in the literature (e.g. Pardo et al. 2013). This poses a problem for research examining predictors of individual differences in imitation: if a specific acoustic feature cannot be used as a proxy, what is the best solution?

One commonly proposed alternative that avoids the issue of feature-specificity is to use perceptual, instead of acoustic, metrics of convergence (see Pardo et al., 2013; Pardo et al., 2017 for discussion). Usually based on independent listeners' judgments of similarity of the imitated speech to the model/exposure speech, these are more holistic, in that these perceptual judgments take into account multiple features simultaneously. However, while one commonly-noted problematic aspect of acoustic metrics is that they rely on specific, researcher-selected dimensions, perceptual metrics also rely on a subset of dimensions, but these are dimensions selected by the listener, and unlike with acoustic metrics, we cannot control – or directly know – what listeners are paying attention to. It is not necessarily the case that listeners are paying attention to all dimensions that are systematically being manipulated by speakers. Therefore, perceptual metrics may fail to capture relevant information about what a speaker is actually doing. While the choice of the best metric will rely on the specific research question, we echo previous work that emphasizes the importance of considering both acoustic and perceptual metrics, as they capture different types of information (Pardo et al., 2013; Pardo et al., 2017), when moving toward a more complete understanding of phonetic imitation.

Another methodological issue is that of statistical power. The current study was designed to be able to capture correlations below r = 0.3, resulting in a larger sample size (n = 133) than that of the previous experimental work on this topic. This is likely the reason for significant results that were not found in previous work: specifically, we found a relationship between individual extents of imitation in implicit vs. explicit tasks, which had not been found in direct tests of the same question by Sato et al. (2013) and Garnier et al. (2013) (n = 12 and n = 14 respectively). It also strengthens our ability to interpret the fact that certain relationships were *not* found in our study, particularly the lack of cross-feature correspondence in implicit imitation. As always is the case with null results, it is not possible to prove that there is no relationship: it could be the case that there is a correlation that was too weak to be detected with our sample size (i.e. a correlation below r = 0.3). However, it is worth noting that we did find significant correlations between *within*-feature correspondences in individual imitation, so it is not the case that there was simply too little power to predict any structured variability at all.

### 4.6. Limitations

One limitation of this study is that while one of its motivations was to test the generalizability of individual performance across features and tasks, we used only two features and two tasks. Clearly, for a more robust test of this question, imitation across a wide range of different features and different tasks need to be considered. For features, it will be important not only to confirm our proposal that speakers show more stability in imitation in more related phones or features, but also to examine the scope of this stability (i.e. identifying the relevant classes, phonetic features, or dimensions, over which individual differences in imitation generalize). In addition, examining features that differ systematically in their phonological patterning and social connotations is necessary to determine the locus of phone- or feature-specific imitation effects.

There are many different types of tasks used to elicit both types of imitation, and the chosen task for a specific study will inevitably impact the results. As discussed in the Methods section, the imitation tasks used in this work were designed with the consideration of keeping the explicit and implicit tasks as similar as possible procedurally (while still maintaining the spirit of the implicit vs. explicit distinction), such that any differ-

ence in performance could be attributed to the implicit vs. explicit nature of the task. Since the two tasks were as comparable as possible, and both features were tested in the same two tasks, we do not expect that our methodological choices have a substantial impact on the patterns of results that are the primary focus of this study: uniformity of individual patterns across features and across implicit vs. explicit imitation. However, here we discuss ways that our choice of task might affect the group-level results, in comparison to methods used in other studies examining phonetic imitation in word-repetition tasks.

For the explicit task, we chose to present the stimuli in minimal pairs in order to direct attention to the target differences. We expect this to increase the overall magnitude of imitation relative to a design where stimuli are presented in a random order (as in e.g. Dufour and Nguyen, 2013). This expectation is based on both perception- and production-based explanations. First, we expect the immediate juxtaposition of the contrasting words will maximize the chance that participants identify the target difference. Second, increasing awareness of minimal pair competitors has been shown to result in contrast enhancement in speech production more generally (e.g. Baese-Berk & Goldrick, 2009), so this would likely apply in the current context as well.

For the implicit task, our participants heard two different versions of tokens varying in a given feature within the same block, which differs from many studies which test how participants respond when exposed to a single phonetic shift across trials (e.g. exposing participants to only lengthened, and not shortened, VOT, e.g. Nielsen 2011). We expect that our paradigm would elicit overall less convergence than one where listeners hear the same type of stimuli across the same block, as convergence effects have been shown to be cumulative, integrating longer-term properties of exposure speech (e.g. Zellou et al., 2017).

As discussed in the introduction, phonetic imitation studies make use of a wide range of methods, and the word repetition tasks chosen for this work are qualitatively very different than some tasks used in other work, including implicit convergence in conversational interaction (e.g. Pardo et al., 2018), "carry-over" effects found in post-exposure production (e.g. Nielsen 2011), or explicit imitation of running speech (e.g. Myers et al., 2024). It will therefore be important in future work to explore the extent to which the findings shown here are general to other types of implicit and explicit tasks.

Finally, given that all three tasks were completed within the same session, it is possible that any structured variation could reflect idiosyncratic properties of the session (e.g. if a participant happens to be sleepy, or was using a different type of headphones than another participant). We do not think that this is a likely explanation for the uniformity found in our study, particularly since uniformity was not found in all cases (e.g. no cross-feature uniformity in implicit imitation), and given the fact that that previous work has found that individual variability is relatively stable across sessions for a given task (e.g. Wade et al., 2021), but testing across multiple sessions and/or in multiple listening environments would be a way to confirm this more robustly.

Although much work is left to be done, we hope that this study has built on previous work which calls for the need to do more systematic testing of the structured variability of individual imitation (e.g. Cohen Priva and Sanker, 2020; Wade et al., 2021; Myers et al., 2024). We think that moving forward with high-powered studies and replications, using multiple metrics to quantify imitation, is the best strategy to tackle the issues of reliability, statistical power, and generalizability cited above.

### 4.7. Conclusion

Understanding the sources and structure of individual variability in imitation is necessary for arriving at accurate models of various linguistic phenomena, including the dynamic nature of speech during conversational interaction, the propagation of sound change, and predictors of success in foreign language learning. Knowing the extent of generalizability of individual patterns is also important from a methodological point of view, as this knowledge helps to gauge the amount of confidence we can have in individual performance metrics from a single task. This work examined the structure of individual variability in English speakers' imitations of subphonemic differences in VOT of voiceless stops and in F2 of the vowels /æ/ and /u/, in controlled implicit and explicit imitation tasks. Our primary goal was to test the extent to which individual proclivity to imitate was related across different phones and across the different tasks. We found that an individual's extent of imitation was correlated across similar types of sounds (how much an individual imitated the VOT of /p/ was related to their extent of imitation of /t/, and imitation of F2 of /æ/ vs. /u/ were also correlated with one another), and this held in both implicit and explicit imitation. Furthermore, in the explicit imitation task, individual extents of imitation were correlated, albeit less strongly, even across phones in different classes (how much an individual imitated VOT of stops was predictive of how much they imitated F2 of vowels). In the implicit task, however, these across-class correlations were not significant. In addition, there was a significant relationship between individual performance on the implicit and explicit task, as well as a role for perception in predicting individual variability in explicit, but not implicit, imitation.

In sum, when considering the question of whether some individuals are overall "greater imitators" than others, the answer appears to depend on what type of imitation is being considered. The results of this work suggest that an individual's extent of explicit imitation may be primarily governed by non-phonetically-selective factors. On the other hand, the extent of spontaneous imitation in a word repetition task depends more heavily on the specific feature being targeted. This suggests a limited role for general cognitive or personality factors in predicting spontaneous imitation: if there is a role, it appears to be too weak to be detectable, even in a highly controlled design. Nevertheless, individual performance in spontaneous convergence is not random, but rather characterized by structured variation, as evidenced by the within-class correlations in extent of implicit imitation. We hope that this work provides a jumping-off point for work examining the scope and

limits of this individual uniformity, as well as more focused tests of factors predicting individual variability in phonetic imitation, in its various forms.

### CRediT authorship contribution statement

**Jessamyn Schertz:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

### Funding

This research was supported by the National Sciences and Engineering Research Council of Canada.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### References

Aguilar, L., Downey, G., Krauss, R., Pardo, J., 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.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x.

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.

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

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. (2015). *Lme4: Linear mixed-effects models using eigen and S4*. https://cran.r-project.org/web/packages/lme4/index.html.

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer, version 6.1.56*. http://www.praat.org.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47. https://doi.org/10.1016/j.wocn.2017.01.001.

Chodroff, E., & Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language, 98*(2), 250–289. https://doi.org/10.1353/lan.2022.0007.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America, 144*(3), EL172–EL177.

Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America, 147*(1), 671–683. https://doi.org/10.1121/10.0000555.

Cohen Priva, U., & Sanker, C. (2020). Natural leaders: some interlocutors elicit greater convergence across conversations and across characteristics. *Cognitive Science, 44*(10). https://doi.org/10.1111/cogs.12897.

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.

Coumel, M., Christiner, M., & Reiterer, S. M. (2019). Second language accent faking ability depends on musical abilities, not on working memory. *Frontiers in Psychology, 10*(257). https://doi.org/10.3389/fpsyg.2019.00257.

De Rosario-Martinez, H. R. (2015). *Package "phia."* https://CRAN.R-project.org/package=phia.

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

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

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology, 4*, 600. https://doi.org/10.3389/fpsyg.2013.00600.

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

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics, 49*(2), 187–200.

Henderson, A., & Rojczyk, A. (2023). Foreign-language accent imitation: Matching production with perception. *English Pronunciation Teaching: Theory, Practice, and Research Findings*, 102–117.

Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics, 39*(4), 683–693. https://doi.org/10.1016/j.wocn.2011.07.003.

Jochim, M., Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2024). *emuR: Main Package of the EMU Speech Database Management System*. https://CRAN.R-project.org/package=emuR.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1594. https://doi.org/10.1037/xhp0000410.

Keshet, J., Sonderegger, M., & Knowles, T. (2014). *AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program], version 0.91*.

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience, 34*(6), 769–786. https://doi.org/10.1080/23273798.2019.1582787.

Kopečková, R., Wrembel, M., Gut, U., & Balas, A. (2021). Differences in phonological awareness of young L3 learners: an accent mimicry study. *International Journal of Multilingualism, 1–17*. https://doi.org/10.1080/14790718.2021.1897127.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13.

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

Lee, Y., Goldstein, L., Parrell, B., & Byrd, D. (2021). Who converges? Variation reveals individual speaker adaptability. *Speech Communication, 131*, 23–34. https://doi.org/10.1016/j.specom.2021.05.001.

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

Lewandowski, N., & Jilka, M. (2019). Phonetic Convergence, Language Talent, Personality and Attention. *Frontiers in Communication, 4*. https://doi.org/10.3389/fcomm.2019.00018.

Makashay, M. J. (2003). *Individual differences in speech and non-speech perception of frequency and duration* [PhD, The Ohio State University]. https://search.proquest.com/openview/b28ee37ace84390067ceb7c9923e216e/1.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *In Interspeech, 2017*. https://doi.org/10.21437/interspeech.2017-1386.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: evidence from the shadowing task. *Cognition, 109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002.

Mora, J. C., Rochdi, Y., & Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness, 23*(1–2), 57–75.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9*(5–6), 453–467.

Myers, E. B., Olson, H. E., & Scapetis-Tycer, J. (2024). Individual differences in accent imitation. *Open Mind, 8*, 1084–1106. https://doi.org/10.1162/opmi_a_00161.

Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition, 44*(2), 580–605. https://doi.org/10.1017/S0272263121000371.

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

Nielsen, K. Y., & Scarborough, R. (2015). Perceptual asymmetry between greater and lesser vowel nasality and VOT. Proceedings of *ICPhS*, Glasgow.

Olmstead, A. J., Viswanathan, N., Aivar, M. P., & Manuel, S. (2013). Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology, 4*, 475. https://doi.org/10.3389/fpsyg.2013.00475.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology, 4*, 559. https://doi.org/10.3389/fpsyg.2013.00559.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics, 72*(8), 2254–2264. https://doi.org/10.3758/bf03196699.

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., 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.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495.

Pinget, A.-F. (2022). Individual differences in phonetic imitation and their role in sound change. *Phonetica, 79*(5), 425–457. https://doi.org/10.1515/phon-2022-2026.

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Reiterer, S. M., Hu, X., Sumathi, T. A., & Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology, 4* (782). https://doi.org/10.3389/fpsyg.2013.00782.

Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. *Working Papers in Phonetics and Phonology 2015*. https://www.researchgate.net/profile/Chelsea-Sanker/publication/276472413_Phonetic_convergence_in_multiple_features/links/5a9034b30f7e9ba4296b7c20/Phonetic-convergence-in-multiple-features.pdf.

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J.-L., & Nguyen, N. (2013). Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology, 4*, 422. https://doi.org/10.3389/fpsyg.2013.00422.

Schertz, J., Adil, F., & Kravchuk, A. (2023). Underpinnings of explicit phonetic imitation: perception, production, and variability. *Glossa Psycholinguistics*. https://doi.org/10.5070/g601123.

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews. Cognitive Science, 11*(2), e1521.

Schertz, J., & Paquette-Smith, M. (2023). Convergence to shortened and lengthened voice onset time in an imitation task. *JASA Express Letters, 3*(2). https://doi.org/10.1121/10.0017066 025201.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422–429.

Spinu, L. E., Hwang, J., & Lohmann, R. (2018). Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English. *International Journal of Bilingualism, 22* (3), 350–370. https://doi.org/10.1177/1367006916681080.

Spinu, L. E., Hwang, J., Pincus, N., & Vasilita, M. (2020). Exploring the use of an artificial accent of English to assess phonetic learning in monolingual and bilingual speakers. *Proceedings of Interspeech, 2020*. https://doi.org/10.21437/interspeech.2020-2783.

Symons, A. E., Holt, L. L., & Tierney, A. (2023). Individual differences in the effects of informational masking on segmental and suprasegmental speech categorization. doi: 10.31234/osf.io/9eq6b.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: contextual influences. *The Journal of the Acoustical Society of America, 125*(6), 3974–3982. https://doi.org/10.1121/1.3106131.

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

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

Ulusahin, O., Bosker, H., McQueen, J., & Meyer, A. (2023). No evidence for convergence to sub-phonemic F2 shifts in shadowing. *Proceedings of ICPhS*. https://pure.mpg.de/rest/items/item_3507211_3/component/file_3507212/content.

Wade, L., Lai, W., & Tamminga, M. (2021). The reliability of individual differences in VOT imitation. *Language and Speech, 64*(3), 576–593. https://doi.org/10.1177/0023830920947769.

Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 297–302). https://doi.org/10.18653/v1/N18-2048.

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.

Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Applied Linguistics, 5*(1), 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815.

Zellou, G., & Brotherton, C. (2021). Phonetic imitation of multidimensional acoustic variation of the nasal split short-a system. *Speech Communication, 135*, 54–65. https://doi.org/10.1016/j.specom.2021.10.005.

Zellou, G., Dahan, D., & Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience, 32*(6), 776–791. https://doi.org/10.1080/23273798.2016.1275710.