

**Running head:** IMITATION OF SECOND LANGUAGE SOUND CONTRASTS

**Full title:** Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts

**Author names:** Miquel Llompart & Eva Reinisch

**Affiliation:** Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich

**Acknowledgements**

This project was funded by a grant from the German Research Foundation (DFG; grant nr. RE 3047/1-1) to the second author. This work is part of the first author's Ph.D. project. We would like to thank Rosa Franzke for her help with testing participants, Jonathan Harrington for advice concerning statistics and Jessica Siddins for comments on a previous version of the manuscript.

**Address for correspondence:**

Miquel Llompart, Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Schellingstraße 3, 80799 Munich, Germany.  
M.Llompart@phonetik.uni-muenchen.de

**Conflict of interest:**

The authors declare no conflict of interest.

This paper has been accepted for publication at *Language and Speech* on 7 September 2018  
The final Journal version of the paper may contain slight changes relative to this preprint.

## **Abstract**

This study investigated the relationship between imitation and both the perception and production abilities of second language (L2) learners for two non-native contrasts differing in their expected degree of difficulty. German learners of English were tested on perceptual categorization, imitation and a word reading task for the difficult English /ɛ/-/æ/ contrast, which tends not to be well encoded in the learners' phonological inventories, and the easy, near-native /i/-/ɪ/ contrast. As expected, within-task comparisons between contrasts revealed more robust perception and better differentiation during production for /i/-/ɪ/ than /ɛ/-/æ/. Imitation also followed this pattern, suggesting that imitation is modulated by the phonological encoding of L2 categories. Moreover, learners' ability to imitate /ɛ/ and /æ/ was related to their perception of that contrast, confirming a tight perception-production link at the phonological level for difficult L2 sound contrasts. However, no relationship was observed between acoustic measures for imitated and read-aloud tokens of /ɛ/ and /æ/. This dissociation is mostly attributed to the influence of inaccurate non-native lexical representations in the word reading task. We conclude that imitation is strongly related to the phonological representation of L2 sound contrasts, but does not need to reflect the learners' productive usage of such non-native distinctions.

*Keywords:* second language learning, imitation, speech production, speech perception, perception-production link.

## **Introduction**

A relevant concern in the study of language is the characterization of the relationship between the speech perception and speech production systems. This link has received considerable attention with regard to native language (L1) listening and speaking, with results often suggesting that the two modalities are tightly connected (Babel, 2009; Goldinger, 1998; Goldinger & Azuma, 2004; Houde & Jordan, 1998, 2002; Lombard, 1911; Meyer, Huettig & Levelt, 2016; Nooteboom & Quené, 2013; Pardo, 2006; Rochet-Capellan & Ostry, 2011; Scott, McGettigan & Eisner, 2009; Wilson, Saygin, Sereno & Iacoboni, 2004). However, it appears less clear how the two systems interact in the course of second language (L2) learning. When learning an L2, both perception and production abilities need to develop so as to allow for the acquisition of a new language system. One aspect that is of special interest concerning this relationship is how perception and production interact throughout the acquisition of the L2 sound inventory, which, unlike the phonological system of the native language, is yet to be fully-established.

There are several ways in which the connection between L2 perception and production during L2 sound learning has been examined. Some studies tested learner populations on their perception and production of specific L2 sounds that are typically challenging for them to learn and then assessed whether performance in one modality is related to performance in the other modality. (Bent, 2005; De Jong, Hao & Park, 2009; Peperkamp & Bouchon, 2011). Others investigated whether improvements following training in one modality transferred to the other modality and, if so, under what circumstances (Akahane-Yamada, McDermott, Adachi, Kawahara, & Pruitt, 1998; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Hirata, 2004; Lopez-Soto & Kewley-Port, 2009;

Thorin, Sadakata, Desain & McQueen, 2018; Wong, 2013). Finally, imitation after a native speaker model has frequently been used as a window into the link between L2 perception and production (De Jong et al., 2009; Flege & Eefting, 1988; Hao & De Jong, 2016; Jia, Strange, Wu, Collado, & Guan, 2006; Rojczyk, 2013; Rojczyk et al., 2013; Schouten, 1977). Imitation tasks are considered to be especially informative regarding this issue because they involve the engagement of the two systems in one task.

There are two main types of paradigms used to elicit speech imitation. On the one hand, shadowing tasks represent an implicit approach because listeners are asked to repeat auditory stimuli as quickly as possible but without instruction to mimic the model speaker. The degree of imitation is then assessed by comparing participants' baseline productions before the task to these repetitions (Fowler, Brown, Sabadini, & Weihing, 2003; Goldinger, 1998; Goldinger & Azuma, 2004; Mitterer & Ernestus, 2008; Nielsen, 2011; Shockley, Sabadini & Fowler, 2004). On the other hand, in explicit imitation tasks, listeners are encouraged to sound as similar as possible to the model speaker they hear when producing the stimuli (Alivuotila, Hakokari, Savela, Happonen, & Aaltonen, 2007; Hao & de Jong, 2016; Kent, 1973, 1974; Kent & Forner, 1977; Repp & Williams, 1985; Rojczyk, 2013; Rojczyk et al., 2013; Zajac & Rojczyk, 2014). In the present study, we used the latter approach to examine L2 learners' imitation, and crucially investigated how imitation performance relates to performance in other tasks in which each of the two subcomponents involved in imitation are probed in isolation (i.e., phonetic categorization in perception and word reading without a model speaker in production).

Perception, imitation (i.e., perception + production) and production by German learners of English were assessed for two English sound contrasts (/i/-/ɪ/ and /ɛ/-/æ/) that differed in how difficult they were expected to be for this learner population, as predicted by models of L2

acquisition (Best & Tyler, 2007; Flege, 1995) and based on previous literature examining the acquisition of these sounds (Bohn & Flege, 1990, 1992; Eger & Reinisch, 2017, 2018; Iverson & Evans, 2007, Llompart & Reinisch, 2017, in press). On the one hand, German-speaking learners of English typically master the English /i/-/ɪ/ distinction because German has a vowel contrast that closely matches the English contrast articulatorily and acoustically (Jongman, Fourakis, & Sereno, 1989; Moulton, 1962; Strange, Bohn, Trent, & Nishi, 2004). This is why this contrast will be henceforth referred to as “easy” L2 contrast. On the other hand, /ɛ/-/æ/ is challenging for native speakers of German because /æ/ is not part of the German sound inventory and, at least at the initial stages of L2 learning, both English /ɛ/ and /æ/ tend to be perceived and produced as the German category /ɛ/, which is similar to English /ɛ/ (Bohn & Flege, 1990; Eger & Reinisch, 2017, 2018; Flege, Bohn, & Jang, 1997, Llompart & Reinisch, 2017, in press). Therefore, for accurate perception and production, learners need to establish /æ/ as a new L2 phonological category that is different from English (and German) /ɛ/. For these reasons, /ɛ/-/æ/ will be henceforth referred to as “difficult” L2 contrast.

In the present study, we asked two main research questions, each of them associated with a specific relationship between imitation and one of its subcomponents, that is, perception and production. First, the relationship between imitation patterns and perceptual categorization was assessed in order to test whether the imitation of L2 sound contrasts is reliably linked to how accurately the two sounds are distinguished in perception. Importantly, following previous research on L2 learning (Pallier, Bosch, & Sebastián Gallés, 1997; Amengual, 2016), we take perceptual abilities as an indicator of how robustly the phonological categories corresponding to the two sounds in the contrast are encoded and differentiated from one another. In this way, the relationship between imitation and perception will inform us on whether imitation patterns are

modulated by the way L2 contrasts are represented in the non-native phonological system of the learners. Secondly, we investigated the relationship between imitation and production abilities so as to determine to what extent imitation of L2 contrasts resembles the learners' productive usage of these contrasts in L2 words. These relationships were examined for an easy vs. a difficult L2 contrast because (i) the two contrasts differ in how well-established they should be in the L2 phonological inventory of the learners (thus related to the first research question), and (ii) they also differ in how likely they are to trigger difficulties at the word level (thus related to the second research question).

The first question that this study set out to address was to what extent L2 imitation is constrained by the degree to which the categories to be imitated are differentiated/established in the imitator's phonological system. There are several findings that suggest that phonological categories exert a strong influence on imitation patterns, that is, that speakers tend to imitate stimuli in accordance with the representations they have for the sounds in their phonological inventory. This is the case in both implicit and explicit imitation tasks (Alivuotila et al., 2007; Flege & Eefting, 1988; Jia et al., 2006; Mitterer & Ernestus, 2008; Nielsen, 2011; Schouten, 1977). Nielsen (2011) showed that native speakers of English did not imitate /p/-initial words with shortened Voice Onset Times (VOTs; the time between the release of a stop closure and the start of vocal fold vibration) for /p/, because accurate imitation of these tokens could introduce ambiguity about the identity of the first consonant of the words. That is, imitation was constrained by the existence of a phonological category /b/ with VOTs similar to those of the /p/-stimuli with shortened VOTs.

Another type of evidence that imitation is constrained by phonological categories is that native speakers have been shown to be unable to accurately imitate sound continua in which fine

phonetic detail changed gradually/in many small steps (Alivuotila et al., 2007; Kent, 1973; Repp & Williams, 1985; Schouten, 1977). Alivuotila et al. (2007) showed that Finnish children as well as adults imitated steps along an /æ/ to /ɑ/ continuum with an abrupt shift in second formant (F2) frequencies from values representative of one category to F2 values typical of the other category (children even more so than adults). That is, they always imitated the different continuum steps as instances of one of their two native phonological categories /æ/ or /ɑ/. Only phonetically-trained individuals who were used to listening to fine phonetic differences were able to closely imitate the intermediate steps of the continuum.

Research on L2 learning has also reported phonological mediation in imitation (Flege & Eefting, 1988; Jia et al., 2006; Schouten, 1977). Schouten (1977) tested L1-Dutch L2-English bilinguals' perception and imitation of a vowel continuum ranging from /i/ to /æ/. When asked to perceptually categorize these sounds as either their L1 or L2 categories, participants grouped the steps quite systematically into discrete categories. Critically, when they were asked to imitate the same continuum steps, imitation patterns closely matched the categorization clusters in the perception task, rather than matching the acoustic properties of the imitated continuum steps. Performance in imitation therefore aligned with perceptual categorization, suggesting that imitation patterns were constrained by the speakers' L1 and L2 phonological categories.

Along the same lines, Flege and Eefting (1988) tested the perception and imitation of a VOT continuum by three different populations: Spanish monolinguals, Spanish speakers of English and English monolingual speakers. They found that both groups of monolingual speakers imitated the continuum according to the phonological categories of their respective native language. That is, Spanish monolinguals produced only lead and short-lag VOTs, while English monolinguals only produced short-lag and long-lag VOTs. By contrast, native Spanish

speakers who also spoke English produced VOTs falling into all three clusters (i.e., lead, short-lag, and long-lag) corresponding to the combination of their L1 and L2 categories. Importantly, similar to the study by Schouten (1977), shifts in produced VOTs in imitation closely matched the listeners' boundary locations in a perceptual identification task.

Finally, Jia et al. (2006) investigated the perception and imitation of several American English vowel contrasts (e.g., /i/-/ɪ/, /ɛ/-/æ/, /æ/-/ɑ/) by native speakers of Mandarin Chinese. Perception was assessed by means of an AXB discrimination task on natural native productions of these vowels, and the same stimuli were also used in an imitation task. Results showed that accuracy in the AXB task over all contrasts correlated with overall accuracy in vowel imitation as determined by native speakers' judgements, again suggesting that imitation is conditioned by how accurately the sounds in the contrast are distinguished perceptually. However, Jia et al. (2006) did not examine this relationship separately for the different sound contrasts they tested. Since not all L2 sound contrasts may be equally problematic for non-native speakers, a question that remains unanswered is whether the apparent connection between perception and imitation is observable for all types of sound contrasts - those that are easy and those that are typically difficult to master in the L2. Differences may be expected because, for easy L2 contrasts, categories should be well-defined and stable in both perception and production, while for difficult L2 contrasts perception and production abilities are expected to be in development, that is, learners should still be in the process of establishing these sounds as two separate L2 phonological categories.

Despite the abundant evidence speaking for strong phonological mediation in imitation, there are also several findings contradicting this view. A case in point are studies showing that speakers can imitate phonetic detail that is not phonologically relevant for them. Therefore, the

acoustic properties of imitated tokens do not always match with how speakers would produce the target categories in the absence of an auditory model (Fowler et al., 2003; Goldinger, 1998; Hao & de Jong, 2016; Rojczyk, 2013; Rojczyk, Porzuczek, & Bergier, 2013; Shockley et al., 2004; Zajac & Rojczyk, 2014). For example, Fowler et al. (2003) and Shockley et al. (2004) showed that native speakers of English could imitate sub-phonemic differences in their native language. Listeners were found to produce longer VOTs for voiceless stop consonants when imitating stimuli with VOTs that had been lengthened artificially, in spite of the fact that these VOT values are not usual in natural productions of English voiceless stops (note that, contrary to Nielsen, 2011, here no ambiguity with another category was created).

Another line of evidence taken to speak against imitation being phonologically constrained is research on the imitation of difficult L2 sounds (Hao & de Jong, 2016; Rojczyk, 2013; Rojczyk, Porzuczek, & Bergier, 2013; Zajac & Rojczyk, 2014). Second language learners have been found to produce sounds that are typically challenging for them in perception and production in a more native (i.e., target)-like manner when imitating a native model than when producing them without a native auditory prompt such as in a reading task. For instance, Rojczyk (2013) found that Polish learners of English produced the non-native vowel /æ/ in a more native-like manner when they imitated English words produced by a native model than when they produced the same words in a reading task without a model speaker. In a similar way, Hao and de Jong (2016) showed that productions of Mandarin tones by English learners of Mandarin were rated as more accurate by native speakers of Mandarin if produced as part of an imitation task than the same speakers' productions of these tones in a read-aloud task. Building on these findings, Hao and de Jong (2016) attribute the better performance in imitation than reading to the fact that phonological categorization is bypassed during imitation (see also Chistovich, Fant,

deSerpa-Leitao & Tjerlund, 1966). According to Hao and De Jong, this by-passing of phonological categorization allows learners to faithfully imitate sounds for which they do not have a native-like phonological representation.

In the present study we therefore first addressed to what extent imitation is modulated by phonological categories by assessing the relationship between imitation and perceptual categorization abilities (as a proxy for phonological encoding for the target sounds) at the level of the individual for two L2 contrasts; one that is supposedly easy to acquire and one that is expected to pose considerable difficulties to our learner population. Secondly, we asked whether productions of sounds in L2 contrasts resulting from imitation of a native model relate to how the same categories are productively used when articulating L2 words without exposure to an auditory model. This second research question was further motivated by the fact that one factor that may contribute to this dissociation, and has not been considered in this context, is the involvement of the L2 lexicon when the production/reading task involves real L2 words (e.g., Rojczyk, 2013).

In the L2 literature, it has been shown on many an occasion that learners are more accurate with difficult L2 contrasts in tasks that only assess their phonetic abilities than in tasks in which the retrieval of L2 words from the mental lexicon is involved (Amengual, 2016; Broersma, 2005; Darcy, Daidone, & Kojima, 2013; Díaz, Mitterer, Broersma, Escera, & Sebastián-Gallés, 2016; Díaz, Mitterer, Broersma & Sebastián-Gallés, 2012; Llompарт & Reinisch, in press; Sebastián-Gallés & Baus, 2005; Sebastián-Gallés, Echeverría, & Bosch, 2005; Simonchik & Darcy, 2018). A likely explanation for the enhanced difficulties in tasks involving the L2 lexicon is that, even when learners can already tease the two sounds of an L2 contrast apart in phonetically-oriented tasks, the distinction may still not be well-established in their

lexical representations for L2 words. Particularly, evidence from eye-tracking suggests that, for difficult L2 contrasts like /ɛ/-/æ/ for German learners of English, the L2 category that is a worse fit to the L1 category (i.e., /æ/) is not reliably encoded in L2 words (Cutler, Weber, & Otake, 2006; Escudero, Hayes-Harb & Mitterer, 2008; Llompart & Reinisch, 2017; Weber & Cutler, 2004). Words with /æ/ are therefore “fuzzy” in terms of the vowel represented in them. This lack of phonetic detail in L2 lexical representations adds to any potential confusion at the sound category level and, critically, heightens the uncertainty about which L2 words carry which of the vowels in the contrast. In light of this, in the present study we elicited productions of the sounds in the two L2 contrasts in an explicit imitation task, which was designed to focus on phonetic detail and trigger scarce lexical involvement, and compared these productions to those from a word reading task in which access to L2 lexical representations was more likely to have an influence on the learners’ performance.

### **The present study**

German learners of English were asked to perform three tasks, separately for the /ɛ/-/æ/ and /i/-/ɪ/ contrast: (i) a perceptual categorization task, (ii) a phonetic imitation task, and (iii) a word reading (i.e., production) task. In the categorization task, we examined the learners’ categorization of a continuum spanning the two vowel sounds of each contrast (*bet - bat* and *sheep - ship*). In the imitation task, participants were asked to imitate the tokens of these continua. In the word reading task, learners had to read aloud a list of written words in English. Results from these tasks were used to assess i) group-level differences between the two contrasts within each task (across all participants) and, most importantly, ii) the relationships between perception and imitation, and imitation and word reading, separately for each contrast at the level of the individual (i.e., within participants).

Regarding learners' performance with the two sound contrasts within each task, consistent differences between the easy and the difficult L2 contrasts were expected. We predicted that learners would perceive and produce the difficult /ɛ/-/æ/ contrast less accurately than the easy /i/-/ɪ/ contrast, because /æ/ is likely not yet quite as well established as a new L2 category as the other critical sounds. This means that learners should show a sharper perceptual boundary for /i/-/ɪ/ than for /ɛ/-/æ/ in the categorization task. Similarly, in the word reading task, we predicted that learners would easily produce /i/ and /ɪ/ as two different categories but they would show considerable acoustic overlap between /ɛ/ and /æ/. Finally, for the imitation task, diverging predictions can be made about the two sound contrasts depending on whether phonological encoding is assumed to play a major role in imitation or not. If phonological categories exert an influence on L2 imitation (e.g., Alivuotila et al., 2007; Flege & Eefting, 1988), imitation patterns should be found to be categorical to a certain extent (i.e., abrupt change in F1 and F2 values at a certain location of the continuum). In addition, more categorical imitation should be expected for the continuum corresponding to the easy /i/-/ɪ/ contrast than for the difficult /ɛ/-/æ/ contrast. This prediction is based on the assumption that the two vowels in the /i/-/ɪ/ contrast are well represented in the L2 phonological system of German learners of English, while the phonological encoding of the sounds in the difficult contrast (especially /æ/) is likely to be less robust and therefore less likely to strongly modulate imitation patterns. If, on the contrary, imitation by-passed phonological encoding (e.g., Hao & De Jong, 2016) and relied on phonetic mechanisms, then imitation patterns would be predicted to mostly change gradually along the continuum. In addition, under such an account, no difference needs to be expected between the two sound contrasts, since the phonological representation of the target categories should not be a major determinant of the learners' imitation patterns.

Similarly, regarding relationships between the different tasks in terms of individual performances, diverging predictions can be made about the relationship between perception and imitation on the one hand, and imitation and word reading on the other hand, depending on the role attributed to phonological encoding in imitation. Specifically, if imitation performance with L2 sound contrasts were indeed tightly related to the phonological encoding of these sounds, we would expect to find a within-individual relationship between responses to the different steps of the continua in perceptual categorization and the acoustic properties of imitations of these same steps. Learners with a more clear-cut (i.e., more categorical) perceptual differentiation of the difficult contrast would be expected to also show a more categorical behavior in imitation, that is, they should produce a more abrupt transition from one category to the other. Learners with less well-defined perceptual boundaries should in turn show less categorical imitation patterns. This should be observable especially for the /ε/-/æ/ contrast, where, due to the difficulty of the contrast, individual variation is expected to be large. For /i/-/ɪ/, it is possible that a relationship between perceptual categorization and imitation cannot be captured at the level of the individual because German learners of English should all have similarly sharp perceptual boundaries and possibly also display similarly categorical imitation patterns. By contrast, if phonological encoding was by-passed in imitation (Chistovich et al., 1966; Hao & De Jong, 2016), then there is no a priori reason to predict a relationship at the level of the individual between perceptual categorization and imitation of a sound continuum for any of the two contrasts.

Concerning the relationship between imitation and production in word reading, there are at least two potential reasons to expect that the acoustic characteristics of imitations of prototypical /ε/ and /æ/ tokens (i.e., the endpoints of the continuum) and read aloud tokens of the same vowels will differ. Firstly, previous studies indicate that L2 learners are frequently better at

imitating difficult L2 categories than at producing them in a word-reading task without an auditory model (Hao & de Jong, 2016; Rojczyk, 2013; Rojczyk et al., 2013). If this was the case, /ɛ/ and /æ/ tokens produced in imitation may not match those produced in word reading. Secondly, it is quite likely that the word reading task used to assess production involves a greater reliance on the L2 lexicon than the imitation task. Since the weak encoding of difficult L2 sounds to lexical representations generally adds to any difficulties learners may have with these sounds at the phonological category level (e.g., Broersma, 2005; Darcy et al., 2013), the relatively higher lexical involvement in the word reading task may have additional negative consequences on learners' production in that task. While these predictions are expected to hold at the group level, it is less clear what the pattern will be within each learner. This is because the expected between-task differences may or may not show systematicity within participants.

Finally, for the easy L2 contrast, learners should not have perceptual, production or lexical difficulties with these sounds. Therefore, assuming that learners have well-encoded representations for these vowels at the phonological and lexical level, /i/ and /ɪ/ are predicted to be imitated (as continuum endpoints) and produced in word reading in a reasonably similar way, making the finding of relationship within individuals highly likely. Nonetheless, if learners do not rely on phonological representations to the same extent in the two tasks, that is, if phonological representations play less of a role in imitation than word reading (as Hao and De Jong, 2016 might propose), a relationship at the individual level may not surface.

## **Methods**

### **Participants**

Forty-two native speakers of German (21 females; mean age = 24.93, sd = 4.3) took part in this study in exchange for a small payment. All participants were students at the University of

Munich. None reported hearing problems. Recruitment criteria were that participants (i) had not learned any language other than German before starting to learn English at school, which was at an average age of 9.61 years (sd = 2.21, range = 4–12), (ii) had not spent more than 6 months in an English-speaking country, and (iii) did not study anything related to language and/or linguistics at the university level. The data from two participants had to be excluded because they did not meet these criteria. Participants filled in a background questionnaire assessing a number of self-estimated measures of English proficiency. On a scale from 1 (native) to 7 (very poor) their average evaluation of their overall proficiency was 2.67 (sd = 0.96) and the evaluation of their abilities to comprehend and speak in English was 2.28 (sd = 0.86) and 3.15 (sd = 1.11), respectively.

### **Materials**

Two English minimal word pairs, one for each contrast, were selected to be recorded for the categorization and imitation tasks. The minimal pair for /æ/-/ɛ/ was *bet-bat* and for /i/-/ɪ/ it was *sheep-ship*. The minimal pairs were recorded in a sound-attenuated booth by a 26-year-old native speaker of Southern Standard British English who lived in London before moving to Munich at the age of 22. Each word was recorded multiple times to select tokens such that the two words in each pair were as similar as possible in recording quality, *fundamental frequency* ( $f_0$ ) contour and perceived speech rate. However, since it appeared that the vowels in *bet* and *bat* were consistently produced with some creakiness, we decided to replace the words' onsets (/b/ and vowel) with recordings from tokens of *bed* and *bad* that the same speaker recorded for an unrelated task. These tokens were less creaky. Critically, the vowel duration was shortened to match the original durations of *bet* and *bat* from which the final stops were taken. The reason to keep the final voiceless consonant, rather than using *bed* and *bad* instead, was to avoid the

influence of other sounds that may be difficult to perceive for our learner population. Word-finally voiced obstruents have been shown to be problematic for German learners of English (Eger & Reinisch, 2017; Smith, Hayes-Harb, Bruss & Harker, 2009) because in German word-final obstruents are canonically devoiced (Fourakis & Iverson, 1984; Port & O'Dell, 1985). No splicing was necessary for the selected tokens of *sheep* and *ship*.

Two 21-step continua, one between *bet* and *bat* and the other between *sheep* and *ship*, were created through duration manipulation and formant shifting by means of a Praat script (Boersma & Weenink, 2010). Vowel duration and first and second formant values (F1 and F2; in Hertz) for the endpoints were taken from the naturally-produced tokens, and continuum steps were set to change linearly in all three dimensions. Table 1 shows the endpoint values.

<Insert Table 1 about here>

For the word reading task, 13 English words for each critical sound (i.e., /ɛ/, /æ/, /i/, /ɪ/) were selected to be read by the participants. Words were mono-, di- and trisyllabic and the number of syllables of the target words varied similarly for each sound so as to obtain a representative sample of how the vowels are produced in the different types of words. Importantly, the critical vowel was in stressed position for all words and none of the words resulted in a minimal pair if the vowel was replaced by the other member of the contrast (e.g., \*h[ɛ]ppy), except for two words with /i/ (i.e., *field* and *feeling*). The words to be read were highly frequent, with lexical frequency being similar for the four critical sounds according to the Zipf-scale measure provided by Subtlex-UK (/ɛ/ = 4.66, /æ/ = 4.82, /i/ = 4.61, /ɪ/ = 4.41; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Appendix A shows the full list of words.

## **Procedure**

All participants completed the three tasks in two experimental sessions that were separated by at least one and no more than three weeks. The majority of participants came for the second session exactly one week after the first. Two sessions were required because in the first session participants took part in the categorization task and additionally completed two unrelated tasks for a different project. In the second session, they first completed the word reading task followed by the imitation task. This order was to avoid any potential influences from imitation onto word reading. For the sake of consistency with our research questions and clarity in the discussion of procedure and results, we present the tasks in the order categorization (i.e., perception only) – imitation (i.e., perception + production) – word reading (i.e., production only). Participants were tested individually in a sound-attenuated booth. The categorization task was run on a laptop computer using Psychopy2 software (v.1.83.01; Peirce, 2007). The imitation task and the word reading task were conducted using a desktop computer running *SpeechRecorder* software (Draxler & Jänsch, 2004). Auditory stimuli in the categorization and imitation tasks were presented binaurally over headphones at a comfortable listening level. Productions in the imitation and word reading tasks were recorded by means of a standing condenser microphone placed at approximately 30 cm from participants. Production data were sampled at 44.1 kHz with 16-bit quantization.

## ***Categorization***

The categorization task was divided into two parts, one for the *sheep-ship* continuum and one for the *bet-bat* continuum, which were run in that order (i.e., easy L2 contrast – difficult L2 contrast). This order was chosen because (i) it allowed participants to familiarize themselves with the task with the less problematic contrast, and (ii) a consistent order was needed across

participants in order to control for potential order effects on individual measures. Listeners received instructions in English that, on every trial, they would hear an English word and they would see two pictures on the screen. Their task was to decide which of the pictures corresponded to the word they had just heard. Participants were instructed to press “1” on a regular computer keyboard when the word they heard corresponded to the picture on the left-hand side of the screen (*sheep* and *bet*, respectively) and “0” when the word was the one depicted in the picture on the right-hand side of the screen (*ship* and *bat*, respectively). Orthographic representations of the words (SHEEP-SHIP and BET-BAT) were provided together with written instructions before the categorization task started but not during the task itself. For each part (i.e., *sheep-ship* and *bet-bat*), the 21 steps of the continuum were presented 10 times in random order with the restriction that all 21 steps had to be presented before any of the continuum steps was repeated. There was no time limit for responses and the presentation of the next trial started 0.8 s after the previous button press. Participants were allowed to take a short break after trials 70 and 140. The categorization task took between 12 and 15 minutes.

### ***Imitation***

For the imitation task, the 21-step continua from *sheep* to *ship* and *bet* to *bat* were shortened to 11 steps by removing every other step (i.e., using only steps 1, 3, 5, etc.; henceforth steps will always be referred to as going from 1 to 11). This was to shorten the task. Participants were told that on every trial, they would hear a sequence of two English words and that their task was to imitate as closely as possible the second of those words once the native speaker had finished talking. They were asked to imitate the second word because the 11 selected steps for each of the two continua were always presented as preceded by one of the endpoint stimuli of the same continuum. The two tokens were always separated by 550 ms. The endpoint stimuli were

presented for two reasons. First, they were used to provide learners with a reference or anchor to be compared with the critical stimulus on every single trial. This was expected to emphasize the perceived differences between the continuum steps, since pretests with a comparable population suggested that learners should be able to identify the endpoints accurately. Secondly, and most importantly, the preceding endpoint stimuli additionally served as control for potential effects of the preceding stimulus on any given trial, since in our case the preceding stimulus was by definition always the same within each block. This allowed to control for trial-by-trial contrast effects in the participants' responses (Diehl, Elman, & McCusker, 1978; Diehl, Kluender, & Parker, 1985; Holt, 2006).

The task consisted of 176 trials and it was again blocked by contrast (88 trials per continuum). The continuum for the easy L2 contrast (*sheep-ship*) was presented first, for the same reasons as in the categorization task. For each contrast, the 11 steps were preceded four times by each of the endpoints. In the first four repetitions they were preceded by the /i/ endpoint and /ɛ/ endpoint (step 1 in each of the two continua), respectively, and in the last four repetitions by the /ɪ/ endpoint and /æ/ endpoint (step 11 in each of the two continua). Within each repetition, the 11 steps were presented in a fully random order.

Before the start of the imitation task, the position and gain of the microphone was adjusted on an individual basis. Participants listened to the prompts over headphones and were explicitly told that they had to wait until the native speaker finished talking before imitating. However, in order to ensure that participants' productions got recorded even if they slightly overlapped with the audio file, recording started 400 ms before the end of the audio file. Participants were given 4 s from the end of the audio file to imitate the second word they heard. The next trial started 1 s after the end of the recording. The timing of the imitation task was

automatic but the experimenter monitored the whole recording session so that it could be paused on participants' request, since they were told that they could take a short break at any point. The imitation task took approximately 20 minutes to complete.

### ***Word reading***

Participants were seated in front of the microphone and were told that they would see English words appear on a computer screen that was located at a comfortable reading distance. They were informed that their task was to wait for a visual signal on the screen (i.e., a green light) and then read the given word aloud. The 52 words (13 for each of the four critical sounds) were presented in a fully random order and recordings were manually timed and controlled by the experimenter. When a word was not recorded properly due to noise (e.g., coughing) or mistakes in word reading (e.g., hesitations), the recording was repeated. The word reading task took approximately from 5 to 7 minutes.

## **Results**

Based on our research questions, we conducted two types of analyses. On the one hand, we looked into group-level patterns of categorization, imitation and word reading comparing the two L2 contrasts under examination (i.e., /i/-/ɪ/ vs. /ɛ/-/æ/). On the other hand, we conducted correlational analyses, separately for each contrast, between measures from the three tasks in order to assess whether relationships between performance in the different tasks arose at an individual level. In order to perform these two types of analyses, we obtained four individual measures per participant and contrast quantifying (i) the slope of the perceptual categorization curve, (ii) the slope of the imitation curve (i.e., of the changes in formant values along the continuum steps), (iii) the acoustic difference between imitated tokens of the two continuum endpoints of each contrast, and (iv) the acoustic difference between read-aloud words containing

the two vowels of each contrast. In the present section, we first discuss how each of the metrics was obtained and the group-level analyses that were run on each of them. Afterwards, the correlational analyses are described in detail. Importantly, note that the calculation of individual metrics per task per contrast has the advantage that it allows for the use of the exact same measures in all the analyses conducted, regardless of whether these are aimed at comparisons between the two contrasts within one specific task or the assessment of individual-level relationships between tasks.

### **Categorization**

The categorization task was used to assess how clear-cut the perceptual identification of the two sounds in each contrast was for each participant. Following previous studies (Pallier et al., 1997; Amengual, 2016), we considered the steepness of the categorization curve as an indicator of how consistently participants differentiated between the two L2 categories and therefore of how well-established the contrast was in the learner's L2 phonological inventory. In order to obtain individual, by-contrast measures of steepness, we calculated the slope of the categorization curve by participant and contrast. Categorization data were submitted to a generalized linear regression model with a logistic linking function (lme4 package 1.1-13 in R version 3.2.2) with Response (0-1) as the categorical dependent variable, an intercept term, and a random-effects structure that included a random slope for Continuum step (1-11) over all possible combinations of Participant and Contrast. From the model, we extracted the slope coefficients for each participant for each contrast (i.e., two values per participant, one for /i/-/ɪ/ and one for /ɛ/-/æ/; the closer values are to 0, the flatter the slope of the categorization function). Two participants were found to have a slope going in the opposite direction to all other participants for the difficult /ɛ/-/æ/ contrast. This could either mean that (i) they were unable to distinguish the two sounds to the extent that they

would not show a curve in the expected direction (i.e., more “bet” responses the more *bet*-like the stimulus) or (ii) they did not do the task properly (e.g., confused the keys despite the visual display). Since we could not be completely certain as to what the reason behind their anomalous performance was, data from these two participants were excluded from all further analyses. One further participant was excluded because the categorization data were missing due to equipment failure. The final dataset thus contained data from 37 German learners of English.

The individual values for the categorization slopes for the /i/-/ɪ/ and /ɛ/-/æ/ continua were subsequently submitted to a linear mixed-effects model with Categorization Slope Coefficient as the dependent variable and Contrast (/i/-/ɪ/ – /ɛ/-/æ/) as fixed factor. Contrast was contrast-coded such that /i/-/ɪ/ was coded as -0.5 and /ɛ/-/æ/ as 0.5. The random-effects structure included a random intercept for participant. P-values were estimated by means of Satterthwaite’s approximation for degrees of freedom using the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017). The model revealed a significant effect of Contrast ( $b = -0.19$ ;  $t = -3.38$ ;  $p < .01$ ). The Categorization Slope Coefficients for the easy /i/-/ɪ/ contrast were higher than those for the difficult /ɛ/-/æ/ contrast, indicating that, as expected, the individual slopes of the categorization functions for /i/-/ɪ/ were in general steeper than those for /ɛ/-/æ/. Table 2 shows the mean and standard deviation for each contrast for the Categorization Slope Coefficients and the other three individual measures used in the analyses of the following subsections. The difference between /i/-/ɪ/ and /ɛ/-/æ/ in Categorization Slope Coefficients can also be observed in

Figure 1, which shows the categorization functions (fitted curves as well as data points per step) for the two contrasts across all participants<sup>1</sup>.

<Insert Table 2 about here>

<Insert Figure 1 about here>

### **Imitation**

Each participant produced 176 tokens in the imitation task. For some participants, some tokens had to be recorded more than once because the recording timed out before the participant responded. These trials were immediately repeated. Trials containing no answer were manually removed from the dataset. Recordings were segmented and phonetically annotated using the

---

<sup>1</sup> An additional analysis was conducted with all responses from all participants (rather than one value per participant) in order to confirm the results of the analysis of the by-participant categorization slopes. A generalized linear-mixed-effects model with a logistic linking function was run with response in the categorization task (1 vs. 0) as dependent variable and Contrast (/i-/ /ɪ/ - /ɛ/-/æ/) and Continuum Step (1–11) as fixed effects, as well as their interaction. The random-effects structure contained a random intercept by participant and slopes for Contrast and Continuum Step over participant. An interaction between Contrast and Continuum Step was found ( $b = 0.17$ ;  $t = 5.77$ ;  $p < .001$ ), confirming thus the reported difference between the two contrasts in the steepness of categorization functions.

Munich Automatic Segmentation System (WebMAuS; Kisler, Schiel, & Sloetjes, 2012; Schiel, 1999). Mean F1 and F2 were measured (LPC 25 ms Gaussian window as implemented by *Praat*; Boersma & Weenink, 2010) on a time window spanning from 25% to 50% of the automatically determined vowel duration<sup>2</sup>. This was the time window that proved to be most stable across all acoustic measures from imitation and word reading. Tokens showing F1 and F2 values three standard deviations above or below the overall mean for each continuum were removed from all analyses. This resulted in the exclusion of 44 recordings (0.65% of the dataset). It was checked that these recordings showed inaccurate measurements due to errors in automatic segmentation or formant tracking, and were not simply unusual non-native productions. In the imitation task, we were interested in obtaining and comparing two properties of the participants' productions: (i) the change in the imitations' acoustics as a function of continuum step (i.e., imitation slope), and (ii) the difference produced between the endpoint stimuli of the continua (i.e., magnitude of difference). Imitation slope was relevant in order to examine how abruptly or gradually participants went from imitating steps of the continuum with formant values corresponding to one category to imitating them with values corresponding to the other category. The magnitude of difference was to assess how differentiated the imitations of the two endpoints (i.e., most prototypical members of a category) were.

---

<sup>2</sup> Vowel duration was measured but not further analyzed because it has been shown that, while native German and native English speakers produce /i/ and /ɪ/ with similar F1 and F2 values, they use duration differently as a cue to the contrast (Bohn & Flege, 1992). This is why we restrict our analysis to the dimension that is used to distinguish between /i/ and /ɪ/ in a similar way in the L1 and L2 of German learners of English.

The imitation slope was to be used in comparison with the participants' performance in the perceptual categorization task, that is, to relate how categorically participants perceived the continua with how categorical they were at imitating them. In order to obtain individual imitation slopes, we first computed the difference score between F2 and F1 (F2-F1) for each token (see Llompart & Reinisch, 2017). F2-F1 scores work similarly for the two sound contrasts, since this metric is expected to be higher for the more *sheep*-like and more *bet*-like continuum steps than for the more *ship*-like and more *bat*-like steps. This is because /i/ and /ɛ/ have lower F1 and higher F2 values than /ɪ/ and /æ/ (see Table 1), respectively. Figure 2 shows the values of F2-F1 difference for the two contrasts for each step of the continuum. As can be seen in the figure, the overall pattern of the imitation data follows a somewhat s-shaped form similar to functions describing perception curves. Since our measure for the production data (F2-F1) is continuous and unbounded by any specific values, in addition to being dependent on the degree that each participant differentiated between the two endpoint stimuli, these data were converted into a binary format (described below) matching the perceptual categorization data. This step was taken in order to make the comparison between the two tasks as fair as possible. Similar procedures can be found in previous studies comparing the perception and production of sound categories (e.g., Harrington, Kleber, & Reubold, 2013).

<Insert Figure 2 about here>

In order to transform the acoustic data into a binary variable, each production was classified as being produced as closer to one or the other endpoint category and the variability around the category boundary determined how flat or steep the imitation curve was (i.e., reflecting how categorically or gradually the continuum was imitated). This was done separately

for each contrast. To assess the category boundary according to which the data should be classified for each participant and contrast, normal distributions were first fit on the F2-F1 difference scores for imitations of the endpoint tokens (i.e., steps 1 and 11). The endpoints were considered the typical productions of these sound categories in imitation. Then F2-F1 values from all other productions were automatically classified based on the (Gaussian) distributions of the endpoint productions. For each production we calculated the probability with which this token was a member of one or the other category and classified each token as belonging to the category with the higher probability. Following this transformation into binary data, the same linear regression model as for the categorization task was applied to assess individual decision boundaries (i.e., the 50% crossover points between the categories) as well as the slopes of the functions. The values of the imitation slopes are thus comparable to the individual slope values obtained in the categorization task<sup>3</sup>.

The second measure obtained from the imitation data was the magnitude of difference between imitations of the two continuum endpoints. It was quantified as the Euclidean distance on the F1 X F2 plane (Harrington, Kleber, & Reubold, 2008; Lindblom, 1986) of the two endpoint stimuli (most *sheep*-like vs. most *ship*-like and most *bet*-like vs. most *bat*-like) on a by-participant basis. Euclidean distances were used in order to reduce the effects of physiological

---

<sup>3</sup> Note that the transformation of acoustic values into binary data may presuppose that imitation patterns should be categorical to a certain extent, thus increasing the likelihood that the imitation data resemble the perception data. However, if this method of analysis indeed favored a relationship between categorization and imitation this would affect all data in the same way, that is, data for all participants and, critically, for the two contrasts examined.

variation due to gender differences and to get individual measures for the difference between imitations of prototypical tokens of /i/ and /ɪ/, on the one hand, and /ɛ/ and /æ/, on the other hand. Importantly, for the native speaker who served as the model for imitation, the Euclidian distances between the endpoints are almost identical for the two contrasts of interest (/i/-/ɪ/ = 445 Hz; /ɛ/-/æ/ = 443 Hz). Therefore, as far as the imitation task is concerned, Euclidean distances allow for a fair comparison of the imitations of continuum endpoints for the two contrasts in spite of the fact that the two contrasts are in different regions of the vowel space.

Two linear mixed-effects models were run on the imitation data in order to probe between-contrast differences in imitation slopes and the acoustic distance between the continuum endpoints. The dependent variables were Imitation Slope Coefficients in the first model and Endpoint Euclidean Distances in the second model. The two models contained Contrast and Anchor Endpoint as fixed factors (i.e., Anchor Endpoint was the stimulus that was always presented before the token to be imitated). Contrast was coded such that /i/-/ɪ/ was coded as -0.5 and /ɛ/-/æ/ as 0.5. Anchor Endpoint was coded such that /i/ and /ɛ/ anchors corresponded to -0.5 and /ɪ/ and /æ/ anchors to 0.5. Anchor Endpoint was added to the analyses of the imitation data in order to ensure that, if an effect of Contrast on any of the two dependent variables was found, it did not differ for imitations after both anchor endpoints. Random effects included a random intercept for participant in each model.

For the model with Imitation Slope Coefficients as the dependent variable, results showed that the effect of Contrast was significant ( $b = -0.31$ ;  $t = -7.25$ ;  $p < .001$ ). Imitation Slope Coefficients for the /i/-/ɪ/ continuum were higher than those for the /ɛ/-/æ/ continuum (see Table 2), which indicates that the individual imitation curves for /i/-/ɪ/ were steeper than those for the

continuum spanning the difficult L2 contrast<sup>4</sup>. This can also be observed in Figure 3, showing the functions of the imitation curves across all participants. Anchor Endpoint, by contrast, did not have a significant effect on Imitation Slope Coefficients ( $b = -0.02$ ;  $t = -0.41$ ;  $p = .69$ ) and its interaction with Contrast was not significant either ( $b = -0.01$ ;  $t = 0.02$ ;  $p = .98$ ). The step of the continuum used as anchor in the imitation task did therefore not modulate the difference between the two continua.

<Insert Figure 3 about here>

The second model, with Endpoint Euclidean Distances as the dependent variable, also rendered a significant effect of Contrast ( $b = -258.40$ ;  $t = -13.24$ ;  $p < .001$ ), indicating that the Euclidean distance between the imitations of the two continuum endpoints was significantly larger for the /i/-/ɪ/ continuum than for the /ɛ/-/æ/ continuum (see Table 2). Even though acoustic

---

<sup>4</sup> As for the categorization data, an additional analysis on all imitation data was conducted so as to confirm the results of the analysis of by-participant imitation slopes. A generalized linear mixed-effects model with a logistic linking function was run with the binomially-transformed values (category A vs. category B) as dependent variable and Contrast (/i/-/ɪ/ – /ɛ/-/æ/), Anchor Endpoint (/i/ and /ɛ/ anchors vs. /ɪ/ and /æ/ anchors) and Continuum Step (1–11) as fixed effects, as well as their interactions. The random-effects structure contained a random intercept by participant and slopes for Contrast, Continuum Step and Anchor Endpoint over participant. An interaction between Contrast and Continuum Step was found ( $b = 1.19$ ;  $t = 9.50$ ;  $p < .001$ ), which confirms the critical difference between the two L2 contrasts.

separation between the two endpoints is visible for the two contrasts in Figure 4 (left panels), learners produced a larger difference during imitation of the endpoints of the easy continuum than when imitating those of the difficult continuum. As in the model above, neither Anchor Endpoint ( $b = 17.80$ ;  $t = 0.91$ ;  $p = .36$ ) nor its interaction with Contrast were significant ( $b = -21.62$ ;  $t = -0.55$ ;  $p = .58$ ). The anchor stimulus that learners heard before each critical token to be imitated had no effect on how big the produced acoustic difference between the endpoints of the continuum was.

<Insert Figure 4 about here>

### **Word reading**

In the word reading task, some words were recorded more than once because of noise or mistakes in word reading. The first author removed all faulty recordings so that there were 52 recordings by participant, 13 for each critical sound. Recordings were segmented and annotated phonetically using WebMAuS (Kisler et al., 2012). For each token, mean values for F1 and F2 were measured on a time window spanning from 25% to 50% of the vowel.

Preliminary analyses showed that there were three words with / $\epsilon$ / in which the vowel was followed by [ɥ] (*helmet*, *helpful* and *smell*). This resulted in substantially lowered F2 values for these words in comparison to the rest of words with that vowel (e.g., *lecture*, *desert*). The other vowel in the contrast, / $\text{æ}$ /, was never followed by [ɥ]. Therefore, in order to avoid confounds between segmental context and acoustic measures of the vowels, the tokens with [ɥ] were removed from the analyses. In addition, tokens showing F1 and F2 values three standard deviations above or below the overall mean for each vowel were also excluded. Again, this was

done so as to discard inaccurate measurements stemming from errors in automatic segmentation or formant tracking. Seventeen recordings were excluded (0.88% of the entire dataset). As with the endpoint productions in the imitation task, we calculated the Euclidean distance on the F1 X F2 plane of the two vowels in each contrast (Harrington et al., 2008; Lindblom, 1986) on a by-participant basis. The words used in this task were also recorded by the native speaker for an unrelated task. His productions of the words showed similar Euclidean Distances between /i/ and /ɪ/ (432 Hz) and between /ɛ/ and /æ/ (380 Hz), but these were not identical like the Euclidean Distances for the endpoint stimuli in the imitation task. This suggests that, even though Euclidean Distances appear to be a reasonable means of comparison between the two contrasts, a small inherent bias for Euclidean distances for /i/-/ɪ/ to be higher than for /ɛ/-/æ/ in the words selected to be read cannot be ruled out.

In order to discern whether there was a difference between the acoustic distinction produced by participants in the two contrasts, individual Euclidean Distances between /i/-/ɪ/ and /ɛ/-/æ/ were submitted to a linear mixed-effects model with the Euclidean Distance for Word Reading as the dependent variable and Contrast (/i/-/ɪ/ – /ɛ/-/æ/) as fixed factor. Contrast was again coded so that /i/-/ɪ/ was -0.5 and /ɛ/-/æ/ was 0.5. The random-effects structure included a random intercept for participant. The model rendered a significant effect of Contrast ( $b = -380.44$ ;  $t = -22.43$ ;  $p < .001$ ). The Euclidean Distance between the two vowels in the word reading task was substantially larger for /i/-/ɪ/ than for /ɛ/-/æ/, as shown in Table 2. Importantly, the difference in Euclidean Distance between the two contrasts as produced by the L2 learners was much larger (380 Hz, see Table 2) than that found in the productions of the native speaker (52 Hz). This indicates that the cause of the considerable divergence between /i/-/ɪ/ and /ɛ/-/æ/

cannot solely be a bias in the inherent acoustic differences between the vowels of the two contrasts.

### **Relationship between categorization and imitation**

In order to assess the relationship between perception and imitation patterns for L2 categories, Pearson's product-moment correlations were calculated between Categorization Slope Coefficients and Imitation Slope Coefficients separately for each of the two contrasts under examination. A correlation between these two metrics would indicate that the steepness of the perceptual boundary between the two categories is related to the steepness of the change in acoustics of imitated tokens along the continuum steps. As shown in Figure 5, left panel, for the easy /i/-/ɪ/ contrast, a small-sized (Cohen, 1992) non-significant correlation was found between individual slopes of the categorization curve and individual slopes of the imitation curve ( $r(35) = .19, p = .26$ ). For the difficult L2 contrast, however, the correlation between the two variables was medium-to-large and statistically significant ( $r(35) = .44, p < .01$ ). As can be seen in Figure 5 (right panel), learners with steeper slopes in the categorization task exhibited steeper slopes in their imitation of the /ɛ/-/æ/ continuum. Note, however, that despite the differences in effect size for the correlations for the two contrasts, direct comparisons between the two contrasts have to be drawn with caution. This is because an additional analysis by means of a linear mixed-effects model with Categorization Slope Coefficients as dependent variable and Contrast and Imitation Slope Coefficients as fixed factors failed to show a significant interaction between the two fixed factors ( $b = 0.30; t = 0.73; p = .47$ ).

<Insert Figure 5 about here>

### **Relationship between imitation and word reading**

Pearson's product-moment correlational analyses were also run in order to examine the relationship between the imitation of endpoint stimuli corresponding to the critical L2 categories and production of the same categories when reading words in the non-native language. Once again, analyses were conducted separately for the two contrasts under examination. For the easy /i/-/ɪ/ contrast, a large significant correlation was found between individual Endpoint (i.e., imitation) Euclidean Distances and Word Reading Euclidean Distances ( $r(35) = .52, p < .001$ ). This indicates that for the easy contrast, also shared with German, the magnitude of the difference between the endpoints produced in the imitation task was strongly related to the magnitude of the difference produced by the L2 learners when reading words in the L2. For the difficult L2 distinction, by contrast, only a small-to-medium-sized correlation that did not reach significance was found between the two measures ( $r(35) = .24, p = .15$ ). An additional analysis by means of a linear mixed-effects model with Word Reading Euclidean Distances as dependent variable and Contrast and Endpoint Euclidean Distances as fixed factors showed that the interaction between the two factors just failed to reach significance according to an alpha criterion of  $p < .05$  ( $b = -0.23; t = -1.71; p = .09$ ). Figure 6 shows the correlation plots between the two variables for the easy (left panel) and difficult L2 contrasts (right panel).

<Insert Figure 6 about here>

A closer inspection of the data gives the impression that the lack of a significant correlation for the difficult L2 contrast may be partially caused by the shrinking of the individual Euclidean Distance values, and relatedly of the between-participant variance for these values, in

the word reading task when compared to the endpoint imitation values. None of the L2 learners produced a Euclidean Distance larger than 250 Hz in the word reading task, while all learners were above this value for the imitation task. This striking disparity suggests that learners generally differentiated the two vowels in the imitation task, even if to different extents, but many of them may have merged / $\epsilon$ / and / $\text{\ae}$ / when reading words aloud. The lower panels of Figure 4 show that this is indeed the case. Productions of the two categories are distinguishable on the F1 X F2 plane in the imitation task (left panel) but appear right on top of each other in the word reading task (right panel). The merger seems to be driven by / $\text{\ae}$ / productions, whose F1 and F2 values in the word reading task approximate those of prototypical / $\epsilon$ / productions in any of the two tasks. This can be appreciated when comparing the average F2-F1 metrics for / $\epsilon$ / and / $\text{\ae}$ / corresponding to the two tasks in Table 3.

<Insert Table 3 about here>

### **Discussion**

The present study examined the perception, imitation and production of English sounds by German learners of English. Participants were tested on a categorization task, an imitation task and a word reading task for two English vowel contrasts that differ in how problematic they are for these learners; that is, a potentially challenging distinction (/ $\epsilon$ /-/ $\text{\ae}$ /) and an easy distinction that is similar to an L1 contrast (/i/-/ɪ/). Individual measures for each task were extracted and used, first, to investigate group-level differences between the two contrasts within each task and second, to assess the relationships between imitation and perception, on the one hand, and imitation and production, on the other hand within each learner. This was done separately for

each contrast. Individual relationships between tasks were of special interest because our main goal was to reach a better understanding of L2 imitation, and, in particular, to clarify the extent to which imitation relates to i) the state of the imitated categories in the learner's L2 phonological system, and ii) the learner's actual usage of the categories in L2 words. The first point was addressed by examining whether imitation patterns relate to perceptual categorization patterns. The second relationship was investigated by comparing the learners' ability to acoustically differentiate L2 contrasts when imitating a native model with their ability to differentiate the same contrasts when reading words aloud in the L2.

Analyses comparing performance for the two contrasts for each task showed that, as expected, learners were better at differentiating the easy /i/-/ɪ/ contrast than the difficult /ɛ/-/æ/ contrast both in perception and production (imitative and non-imitative). In the perception task, categorization of the /i/-/ɪ/ continuum rendered steeper categorization slopes (i.e., higher slope coefficients) than the /ɛ/-/æ/ continuum, indicating that the categories of the easy contrast were more sharply differentiated in perception than the categories from the difficult contrast. In the imitation task, the two measures that were examined pointed in the same direction. The slopes of the imitation curves were significantly steeper for /i/-/ɪ/ than for /ɛ/-/æ/, mirroring the results of the perceptual categorization task. Likewise, the acoustic differences between the two endpoint stimuli of the imitation continuum for the easy contrast (/i/-/ɪ/) were larger than for the difficult contrast (/ɛ/-/æ/). Finally, in the word reading task, learners consistently produced /i/ and /ɪ/ as two well-differentiated categories, while words with /ɛ/ and /æ/, were produced with substantial acoustic overlap.

Regarding the relationship between tasks within learners, correlational analyses showed numerically diverging patterns for the two contrasts of interest. On the one hand, for the

relationship between perception (i.e., Categorization Slope Coefficients) and imitation (i.e., Imitation Slopes Coefficients), a medium-to-large correlation was found for the difficult /ɛ/-/æ/ contrast, while for the easy /i/-/ɪ/ contrast the correlation was only small. On the other hand, for the relationship between imitation (i.e., Endpoint Euclidean Distances) and word reading (i.e., Word Reading Euclidean Distances), only a small-to-medium correlation was found for the difficult L2 contrast, while data for the easy L2 contrast showed a large correlation. Analyses comparing these relationships across contrasts, however, revealed that there were no statistically significant differences between the two contrasts, in spite of the numerical differences observed in effect sizes as indicated by the correlation coefficients. Only a marginally significant difference between contrasts was found for the relationship between imitation and word reading, indicating a tendency for this relationship to be stronger for the easy than for the difficult contrast.

The first aim of the study was to shed light on the role of phonological encoding in the imitation of second language contrasts. Results from both the analysis of imitation patterns comparing the two contrasts and the correlations between perceptual categorization and imitation suggest that the representation of L2 categories in the non-native phonological system exerts a strong influence on imitation performance. Group-level analyses (i.e., across participants) of the imitation data showed that the continuum for the /i/-/ɪ/ contrast was imitated more categorically than the /ɛ/-/æ/ continuum (see Figure 3). Based on predictions of L2 models and previous research, the sounds in the former contrast were assumed to be more strongly represented in L2 phonology than /ɛ/ and /æ/ because of the difficulties this contrast has been shown to cause for native speakers of German. Note that the results of the present study indeed support this assumption, since /i/ and /ɪ/ were more sharply differentiated in perception and more clearly

articulated as two different categories in production (both in imitation and word reading) than /ɛ/ and /æ/. Consequently, the fact that the continuum for the “stronger” contrast from a phonological standpoint was imitated more categorically than that for the “weaker” contrast constitutes evidence that imitation is mediated by phonological encoding, speaking thus against the by-passing of phonological processing during imitation (Chistovich et al., 1966; Fowler et al., 2003; Hao & De Jong, 2016).

Critically, within-participant analyses across tasks point in the same direction.

Correlational analyses show that the individual slopes for categorization and the individual imitation slopes are quite consistently correlated for the difficult L2 contrast. This indicates that there is a relationship between how categorical learners’ distinction of /ɛ/ and /æ/ was in perception, which is taken as a proxy of how robustly differentiated their phonological categories are, and how categorically they imitated the same continuum: the more clear-cut the distinction in perception, the more abrupt the shift from one category to the other in imitation. This therefore replicates findings of previous studies outlining a strong connection between perceptual categorization and imitation in a non-native language (Flege & Eefting, 1988; Jia et al., 2006; Schouten, 1977). By contrast, the correlation between categorization slopes and imitation slopes for the easy (/i/-/ɪ/) contrast was small (though only numerically smaller than for /ɛ/-/æ/). In the Introduction, we predicted that a weak relationship at the individual level for this contrast might be observed if learners were really homogeneous in their performances in the two tasks. Nonetheless, an examination of Figure 5 (left panel) suggests that this is not the case. In fact, substantial variation across a wide range of values can be observed in categorization slopes, while imitation slopes show less variability. A likely explanation for the weak relationship between the two is therefore that differences in how variable learners were in the two tasks

obscured a potential relationship when examining individual learners' values. The enhanced variability in perception (vs. imitation) is likely to have been partly caused by individual differences in how sensitive learners are to within-category variation (McMurray, Tanenhaus & Aslin, 2002; McMurray, Aslin & Tanenhaus, 2008) when categorizing the continuum steps, which would in turn determine their categorization slopes. This sensitivity has been documented to play a major role in perceptual tasks on native sounds distinctions, and English /i/-/ɪ/ would be close to native for German learners (i.e., since the contrast is shared by the L1 and L2).

The second research question we addressed was whether productions of L2 sounds in imitation would match/mismatch how the same categories are used when producing L2 words in the absence of a native model. This was assessed by comparing the acoustics of imitations of the vowels in the two continuum endpoints for each contrast with productions of L2 words containing these same vowels. The relationship between imitation and word reading appears to be very robust for the easy /i/-/ɪ/ contrast, as evidenced by a large correlation, but weak at best for the difficult /ɛ/-/æ/ contrast, where the correlation between measures from the two tasks was small and not significant. These results are hence in agreement with previous studies showing that, for difficult L2 contrasts, imitation and production do not always align (Hao & De Jong, 2016; Rojczyk, 2013; Rojczyk et al. 2013)<sup>5</sup>.

---

<sup>5</sup> This seemingly contradicts the findings of Reiterer and colleagues (Reiterer et al, 2011), who found that the imitation abilities of German speakers in an unfamiliar language (Hindi), correlated with L2 (English) production accuracy. However, there are major methodological differences between our study and Reiterer et al. (2011), like the diverging levels of proficiency of participants in the imitated language (learners vs. naïve listeners), the type of stimuli used

The apparent dissociation between imitation and word reading has been attributed to differences in task affordance. Hao and De Jong (2016) argue that responses to the auditory stimuli in imitation do not engage the phonological system to the same extent as other production tasks, like as for instance word reading, do. However, the strong relationship between imitation and word reading found for /i-/ɪ/ goes against this proposal. If learners had simply imitated either the native speaker's auditory-acoustic patterns (Diehl & Kluender, 1989; Holt & Lotto, 2008) or his gestural programs for the target words (Fowler et al., 2003; Galantucci, Fowler & Turvey, 2006) when they were imitating the steps of the continuum between *sheep* and *ship*, there would be no reason to expect such a strong relationship at the individual level between imitation and word reading. This tight connection can only be understood assuming that learners accessed their phonological representations for /i/ and /ɪ/ when imitating the endpoints of such a shared, almost-native contrast.

In our view, the weak relationship between imitation and word reading for the difficult L2 contrast can be partly explained by lexical difficulties acting on top of phonological processing in the word reading task. In addition to reliance on phonological categories, word reading was expected to trigger lexical involvement because learners most likely recognized the words on the screen, and therefore retrieved their stored lexical representations for these words prior to producing them. This opens the door for the possibility that the production of /æ/-words with acoustic values close to /ɛ/ observed in this task may have a lexical origin. Since L2 learners

---

(words with specific L2 categories vs. whole sentences) and the assessment of imitation and production performance (acoustic measures vs. native-speaker judgments). These differences limit the insights to be gained from comparisons between the two studies considerably.

are often unable to accurately encode difficult L2 sounds to lexical representations even when they can already perceive them quite accurately (e.g., Díaz et al., 2012), it is possible that the difficult-to-learn /æ/ category is not robustly mapped to (some of) the learners' lexical representations for the words that they had to read aloud (e.g., *happy*, *champion*). This resulted in them producing these words with vowels approximating the more native-like /ɛ/ category (i.e., merging the contrast; see also Flege et al. 1997). Importantly, unlike the hypothesis put forward by Hao and de Jong (2016), this account of the acoustic overlap observed in the word reading task can easily accommodate our set of findings on the relation between perception and imitation, which points towards a substantial phonological involvement in imitation. Still, note that it cannot be ruled out that differences between the two tasks other than lexical involvement, like cognitive demands (e.g., Bosker, Reinisch & Sjerps, 2017; Mattys & Wiget, 2011) or the presence vs. absence of minimal pairs (Nelson & Wedel, 2017) also played a role in the relationship between imitation and word reading (albeit for both contrasts examined, not only /ɛ-/ /æ/)<sup>6</sup>.

---

<sup>6</sup> As a reviewer mentioned, productions of /ɛ/ and /æ/ by German learners of English may also be related to how they produce their closer L1 counterpart (i.e., German /ɛ/). Kartushina and Frauenfelder (2014) showed that production accuracy for sounds in difficult L2 contrasts could be predicted by the position in the acoustic space and the amount of variability (i.e., category compactness) for the closer L1 category to the L2 contrast (see also Kartushina, Hervais-Adelman, Frauenfelder & Golestani, 2016). However, in the present study no production data in German were collected. It is hence not possible to assess the potential influence of individual differences in L1 production on our results.

What do our results now tell us about the perception-production link in L2 learning? We take the outcome of the present study to indicate that imitation is a good window to how robust the perception and production of L2 contrasts are at the phonological category level and how closely they relate to each other. When the link between perception and production in imitation was investigated in two tasks that mainly focused the learners' attention on the phonological processing of phonetic detail, a relationship between the two could be captured. Crucially, this relationship was especially robust for the difficult L2 contrast, that is, the contrast that involved the establishment of a new category in the L2 phonological inventory. This suggests that perception and production abilities in the course of L2 learning appear to be tightly coupled (see Schmitz, Díaz, Fernández Rubio, & Sebastián-Gallés, 2018, for similar conclusions). Besides, the finding of such a relationship in analyses of individual values additionally suggests that examining imitation together with perceptual categorization can critically shed light on individual differences in perception and production abilities with challenging non-native contrasts.

However, imitation of /ɛ/ and /æ/ did not consistently relate to how learners produced these sounds when they read L2 words aloud. Results showed that when the focus was taken away from processing the fine phonetic detail of the target categories and placed on the retrieval of words in the reading task, learners were in general unable to produce a reliable distinction between the two sounds (Figure 4, bottom-right panel). Therefore, even if in the word reading task learners were expected to produce the words carefully and were exposed to orthographic information that should help them distinguish the two sounds (i.e., <e> vs. <a>), prompting lexical processing (i.e., access to the stored representations of L2 words) seems to have resulted in a massive shrinkage of the acoustic distinction between /ɛ/ and /æ/.

Interestingly, an additional reflection of the observable mismatch for the difficult L2 contrast between perceptual categorization and imitation (i.e., focus on specific sounds), on the one hand, and word reading (i.e., focus on words), on the other hand, is that only the latter relates to the learners' self-reported proficiency and L2 experience. From the background questionnaires that participants filled out, we calculated individual proficiency/experience scores following the procedure in Eger and Reinisch (2017). These scores encompassed the mean of self-rated speaking and comprehension skills, self-reported frequency of speaking and listening in English and self-estimated German accent when speaking in English. All measures were provided on a 1-to-7 scale in which 1 indicated good skills, frequent use and weak accent, and 7 indicated the opposite. When this measure was compared to performance in the three experimental tasks with /ɛ/-/æ/, only the correlation between proficiency/experience and word reading came out significant ( $r(35) = -.38, p < .05$ ): the better (i.e., the closer to 1) the proficiency/experience score, the bigger the difference produced between /ɛ/ and /æ/ in word reading. We take this to suggest that the learners' subjective assessment of their L2 skills may be closely related to how they can productively use "problematic" non-native distinctions. That is, learners' impressions of their own English proficiency and experience seem to relate to their performance in situations constrained by both phonological and lexical influences, rather than to sheer phonological-level abilities (e.g., perceptual categorization) with the same contrasts.

In sum, the present study contributed to the understanding of L2 imitation and the perception-production link in several ways. First, our results provide consistent evidence that imitation of L2 sounds is constrained by how robustly these sounds are represented in the learner's phonological system. This suggests that, as argued by Flege and Eefting (1988), phonological categorization is involved in imitation and goes against the claims of previous

studies arguing that imitation by-passes phonological encoding (Chistovich et al., 1966; Hao & de Jong, 2016). Secondly, findings for the difficult L2 contrast indicate that perception and production after imitation go hand in hand to a great extent, outlining thus a tight perception-production link at the phonological category level for these sounds. Finally, imitation of the vowels of this contrast was not reliably related to production without an auditory model in a word reading task. We argue that, in our study, this dissociation is to a large extent due to the influence of inaccurate non-native lexical representations in the word reading task. In short, imitation is strongly related to the phonological representation of L2 sounds and contrasts, but does not reflect the actual usage of such sounds when the focus is taken away from the sounds themselves.

## References

- Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., & Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. In *Proceedings of ICSLP-1998*, 1-4, Sydney, Australia.
- Alivuotila, L., Hakokari, J., Savela, J., Happonen, R-P., & Aaltonen, O. (2007). Perception and imitation of Finnish open vowels among children, naïve adults, and trained phoneticians. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 361–364). Saarbrücken, Germany: Saarland University.
- Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, 37(5), 1221–1251.  
<https://doi.org/10.1017/s0142716415000557>
- Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. Ph.D. dissertation. Berkeley, CA: University of California, Berkeley.
- Bent, T. (2005). *Perception and production of non-native prosodic categories*. Ph.D. dissertation. Evanston, IL: Northwestern University.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, pp. 13–34. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.17.07bes>
- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer. Version 5.4.22, <http://www.praat.org/>.

- Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(03), 303–328.  
<https://doi.org/10.1017/s0142716400008912>
- Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(02), 131–158.  
<https://doi.org/10.1017/s0272263100010792>
- Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, *94*, 166–176. doi:10.1016/j.jml.2016.12.002
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985.  
<http://dx.doi.org/10.3758/bf03206911>
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*, 2299–2310.  
<https://doi.org/10.1121/1.418276>
- Broersma, M. (2005). *Phonetic and lexical processing in a second language*. Ph.D. dissertation. Nijmegen, The Netherlands: Radboud University.
- Chistovich, L., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966). Mimicking of synthetic vowels. *Quarterly Progress and Status Report, Speech Transmission Lab, Royal Institute of Technology, Stockholm*, *2*, 1–18.

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi:10.1037/0033-2909.112.1.155
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. <https://doi.org/10.1016/j.wocn.2005.06.002>
- Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. <https://doi.org/10.1075/ml.8.3.06dar>
- De Jong, K., Hao, Y. C., & Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics*, *37*(4), 357–373. <http://dx.doi.org/10.1016/j.wocn.2009.06.001>
- Díaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastián-Gallés, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition*, *19*(05), 955–970. <https://doi.org/10.1017/s1366728915000450>
- Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. <https://doi.org/10.1016/j.lindif.2012.05.005>
- Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 599–609. <https://doi.org/10.1037/0096-1523.4.4.599>

- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*(2), 121–144. [https://doi.org/10.1207/s15326969eco0102\\_2](https://doi.org/10.1207/s15326969eco0102_2)
- Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance, 11*(2), 209–220. <https://doi.org/10.1037/0096-1523.11.2.209>
- Draxler, C., & Jänsch, K. (2004). SpeechRecorder—A universal platform independent multichannel audio recording software. *Proceedings of the fourth international conference on Language Resources and Evaluation* (pp. 559–562). Lisbon, Portugal: Centro Cultural de Belem.
- Eger, N. A., & Reinisch, E. (2017). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition, 1*–22. <https://doi.org/10.1017/s0272263117000377>
- Eger, N. A., & Reinisch, E. (2018). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000599
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345–360. <https://doi.org/10.1016/j.wocn.2007.11.002>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233–277. Timonium, MD: York Press.

- Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470. <https://doi.org/10.1006/jpho.1997.0052>
- Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, 83(2), 729–740. doi:10.1121/1.396115
- Fourakis, M., & Iverson, G. K. (1984). On the “incomplete neutralization” of German final obstruents. *Phonetica*, 41(3), 140–149. <https://doi.org/10.1159/000261720>
- Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396–413. [https://doi.org/10.1016/S0749-596X\(03\)00072-X](https://doi.org/10.1016/S0749-596X(03)00072-X)
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377. <https://doi.org/10.3758/bf03193857>
- 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>
- Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, 11(4), 716–722. <https://doi.org/10.3758/bf03196625>
- Hao, Y.C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, 54, 151–168. <https://doi.org/10.1016/j.wocn.2015.10.003>
- Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The*

*Journal of the Acoustical Society of America*, 123(5), 2825–2835.

<https://doi.org/10.1121/1.2897042>

Harrington, J., Kleber, F., & Reubold, U. (2013). The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. *The Journal of the Acoustical Society of America*, 134(1), 551–561.

<https://doi.org/10.1121/1.4808328>

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17, 357–376. <https://doi.org/10.1080/0958822042000319629>

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America*, 120(5), 2801–2817. <https://doi.org/10.1121/1.2354071>

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, 17(1), 42–46.

<https://doi.org/10.1111/j.1467-8721.2008.00545.x>

Houde, J. F. & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. doi:10.1126/science.279.5354.1213

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech Language and Hearing Research*, 45(2), 295.

doi:10.1044/1092-4388(2002/023)

Iverson, P., & Evans, B. G. (2007). English vowel training with different first-language vowel systems. *The Journal of the Acoustical Society of America*, 121(5), 3072–3072.

<https://doi.org/10.1121/1.4781875>

- Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, *119*(2), 1118-1130.  
<https://doi.org/10.1121/1.2151806>
- Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The acoustic vowel space of modern Greek and German. *Language and Speech*, *32*(3), 221–248.  
<https://doi.org/10.1177/002383098903200303>
- Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*.  
doi:10.3389/fpsyg.2014.01246
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, *57*, 21-39.  
<http://dx.doi.org/10.1016/j.wocn.2016.05.001>
- Kent, R. D. (1973). The imitation of synthetic vowels and some implications for speech memory. *Phonetica*, *28*(1), 1–25. <https://doi.org/10.1159/000259442>
- Kent, R. D. (1974). Auditory-motor formant tracking: A study of speech imitation. *Journal of Speech Language and Hearing Research*, *17*(2), 203–222.  
<https://doi.org/10.1044/jshr.1702.203>
- Kent, R. D., & Forner, L. (1977). A developmental study of speech production: Data on vowel imitation and sentence recitation. *The Journal of the Acoustical Society of America*, *62*(S1), S101–S101. <https://doi.org/10.1121/1.2016015>

- Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via web services: The use case WebMAUS. In J. C. Meister (Ed.), *Proceedings of Digital Humanities Conference 2012* (pp. 30–34). Hamburg, Germany: University of Hamburg.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).  
doi:10.18637/jss.v082.i13
- Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, 13-44.
- Llompert, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 1040–1056. <https://doi.org/10.1037/xhp0000383>
- Llompert, M., & Reinisch, E. (in press). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*.
- Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales Des Maladies De LOreille Et Du Larynx*, *XXXVII*(2), 101–109.
- Lopez-Soto, T., & Kewley-Port, D. (2009). Relation of perception training to production of codas in English as a second language. *The Journal of the Acoustical Society of America*, *125*(4), 2756–2756. <https://doi.org/10.1121/1.4784637>
- Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145–160. doi:10.1016/j.jml.2011.04.004
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42. doi:10.1016/s0010-0277(02)00157-9

- McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. doi:10.1037/a0011747
- Meyer, A. S., Huettig, F., & Levelt, W. J. M. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, *89*, 1–7. doi:10.1016/j.jml.2016.03.002
- 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>
- Moulton, W. G. (1962). *The sounds of English and German*. Chicago: University of Chicago Press.
- Nelson, N. R., & Wedel, A. (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics*, *64*, 51–70. doi:10.1016/j.wocn.2017.01.008
- Nooteboom, S. G., & Quené, H. (2013). Parallels between self-monitoring for speech errors and identification of the misspoken segments. *Journal of Memory and Language*, *69*(3), 417–428. doi:10.1016/j.jml.2013.04.006
- 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>
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. [https://doi.org/10.1016/s0010-0277\(97\)00030-9](https://doi.org/10.1016/s0010-0277(97)00030-9)

- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393. doi:10.1121/1.2178720
- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*, 161-164, Florence, Italy.
- Port, R., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13(4), 455–471.
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., Winkler, S., & Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00271
- Repp, B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. *Speech Communication*, 4(1-3), 105–120. [https://doi.org/10.1016/0167-6393\(85\)90039-1](https://doi.org/10.1016/0167-6393(85)90039-1)
- Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory-motor transformations in speech. *Journal of Neuroscience*, 31(7), 2657–2662. doi:10.1523/jneurosci.6020-10.2011
- Rojczyk, A. (2013). Phonetic imitation of L2 vowels in a rapid shadowing task. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Aug. 2012. (pp. 66-76). Ames, IA: Iowa State University.

- Rojczyk, A., Porzuczek, A., & Bergier, M. (2013). Immediate and distracted imitation in second-language speech: Unreleased plosives in English. *Research in Language, 11*(1).  
<https://doi.org/10.2478/v10015-012-0007-7>
- Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the International Congress of Phonetic Sciences* (pp. 607–610). San Francisco, CA: University of California, Berkeley.
- Schmitz, J., Díaz, B., Fernández Rubio, K., & Sebastian-Galles, N. (2018). Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. *Language, Cognition and Neuroscience, 33*(5), 527–546. <https://doi.org/10.1080/23273798.2017.1390142>
- Schouten, M. E. H. (1977). Imitation of synthetic vowels by bilinguals. *Journal of Phonetics, 5*, 273–283.
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10*(4), 295–302. doi:10.1038/nrn2603
- Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 279–292). Hillsdale, NJ: LEA.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*(2), 240–255. <https://doi.org/10.1016/j.jml.2004.11.001>

- Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66(3), 422–429. doi:10.3758/bf03194890
- Simonchyk, A., & Darcy, I. (2018). The effect of orthography on the lexical encoding of palatalized consonants in L2 Russian. *Language and Speech*. doi:10.1177/0023830918761490
- Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, 37(3), 257–275. doi:10.1016/j.wocn.2009.03.001
- Strange, W., Bohn, O. -S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, 115(4), 1791–1807. <https://doi.org/10.1121/1.1687832>
- Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America*, 144(1), 92–103. doi:10.1121/1.5044415
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1–25. [https://doi.org/10.1016/s0749-596x\(03\)00105-0](https://doi.org/10.1016/s0749-596x(03)00105-0)

- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702.  
doi:10.1038/nn1263
- Wong, J. W. S. (2013). The effects of perceptual and or productive training on the perception and production of English vowels /l/ and /i:/ by Cantonese ESL learners. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, & P. Perrier (Eds.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)* (pp. 2113–2117). Lyon, France: ISCA.
- Zajac, M., & Rojczyk, A. (2014). Imitation of English vowel duration upon exposure to native and non-native speech. *Poznan Studies in Contemporary Linguistics*, 50(4), 495–514.  
<https://doi.org/10.1515/psicl-2014-0025>

## Tables

Table 1

*Duration and F1 and F2 values of endpoint vowels of the two continua used in the categorization and imitation tasks.*

Endpoint	Duration (ms)	F1(Hz)	F2 (Hz)
/ɛ/ (bet)	140	541	1804
/æ/ (bat)	180	841	1477
/i/ (sheep)	235	266	2354
/ɪ/ (ship)	120	440	1944

Table 2

*Mean and standard deviation (in parentheses) for the four individual measures analyzed (Categorization Slope Coefficients, Imitation Slope Coefficients, Endpoint Euclidean Distances (in Hz) and Word Reading Euclidean Distances (in Hz)) for the two contrasts of interest (/i/-/ɪ/ and /ɛ/-/æ/).*

Measure	/i/-/ɪ/	/ɛ/-/æ/
Categorization Slope Coefficients	0.72 (0.36)	0.53 (0.21)
Imitation Slope Coefficients	0.63 (0.20)	0.43 (0.13)
Endpoint Euclidean Distances	618 (143)	359 (114)
Word Reading Euclidean Distances	495 (102)	115 (51)

Table 3

*Mean and standard deviation (in parentheses) of F2-F1 values for the four vowels of interest for the endpoint stimuli in the imitation task and the words read aloud in the word reading task*

Vowel	Imitation	Word Reading
/ɛ/	1180 (194)	1177 (197)
/æ/	708 (191)	1042 (204)
/i/	2047 (221)	2034 (256)
/ɪ/	1299 (202)	1430 (239)

## Figure Captions

*Figure 1.* Categorization functions for /i/-/ɪ/ (in gray) and /ɛ/-/æ/ (in black). Lines connect the fitted values by step by contrast and point symbols indicate the raw means for each step.

*Figure 2.* F2-F1 values for each continuum step of the imitation task for /i/-/ɪ/ (left panel) and /ɛ/-/æ/ (right panel). Boxes and whiskers show variability between participants. Scales have been adjusted to accommodate for the F2-F1 values of each contrast.

*Figure 3.* Imitation functions (for binomially-transformed acoustic data) for /i/-/ɪ/ (in gray) and /ɛ/-/æ/ (in black). Lines connect the fitted values by step by contrast and point symbols indicate the raw means for each step.

*Figure 4.* Mean F1 and F2 values for /i/-/ɪ/ (upper panels) and /ɛ/-/æ/ (lower panels) in the imitation task (endpoints of continuum; left) and the word reading task (right). Each symbol represents the mean value for one participant for the corresponding vowel. X- and y-axes show the same frequency range for the two contrasts but differ in absolute values to accommodate for the fact that /i/-/ɪ/ and /ɛ/-/æ/ are in different regions of the vowel space.

*Figure 5.* Correlation between Categorization Slope Coefficients and Imitation Slope Coefficients for /i/-/ɪ/ (left panel) and /ɛ/-/æ/ (right panel). Scales were adjusted so that x- and y-axes are consistent within each panel but they differ for the two contrasts because of their differences in the range of values for Categorization and Imitation Slope coefficients.

*Figure 6.* Correlation between (Imitation) Endpoint Euclidean Distance and Word Reading Euclidean Distance for /i/-/ɪ/ (left panel) and /ɛ/-/æ/ (right panel). Scales were adjusted so that x- and y-axes are consistent within each panel but they differ for the two contrasts because of their differences in acoustic distance between the two vowels.

## Figures

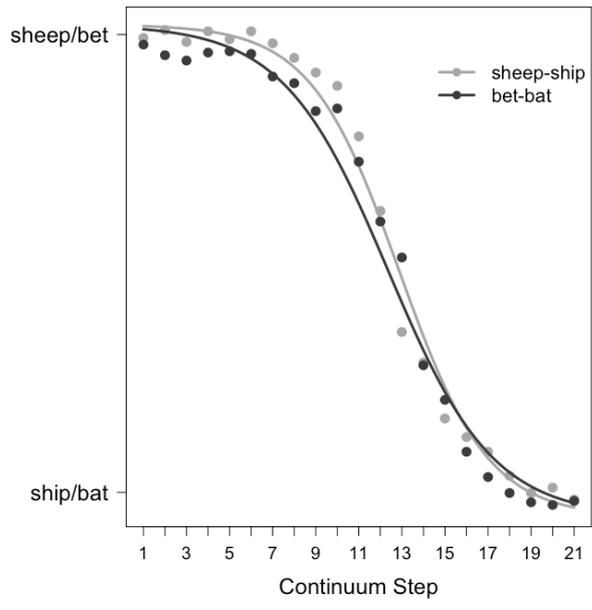


Figure 1

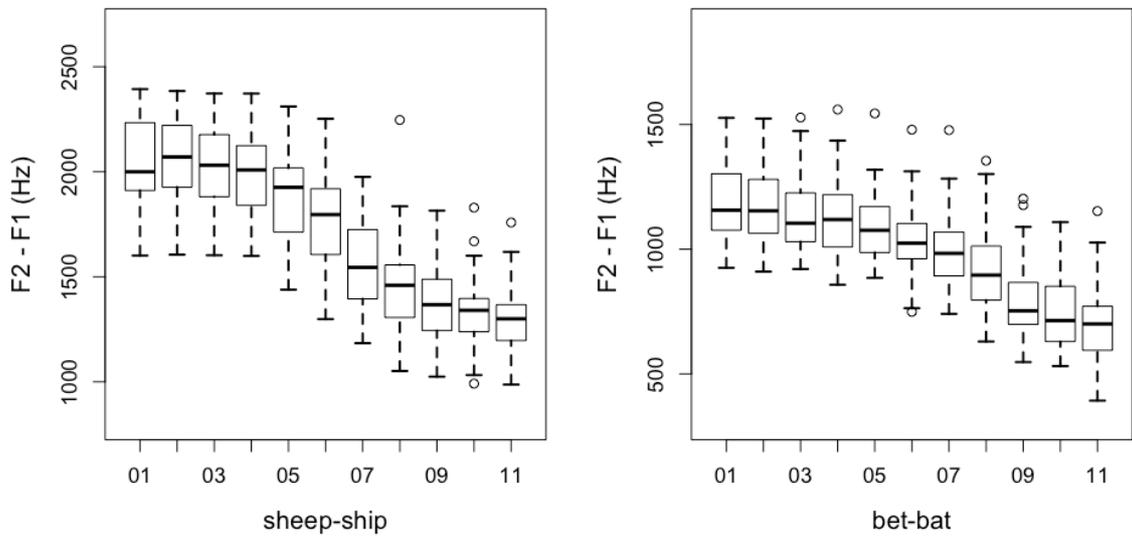


Figure 2

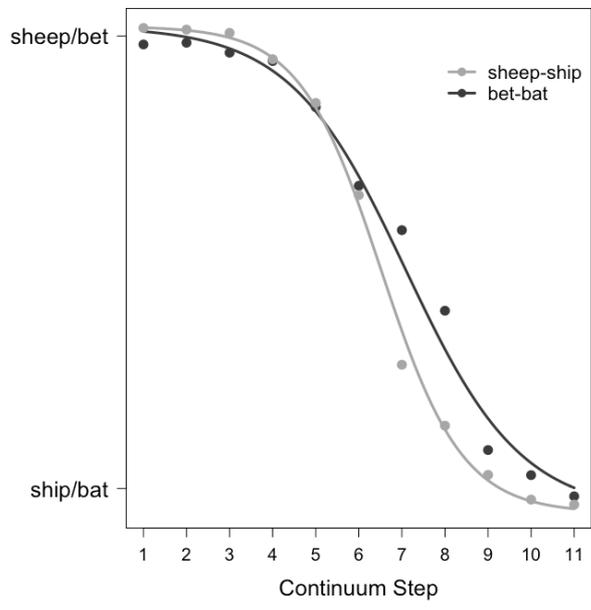


Figure 3

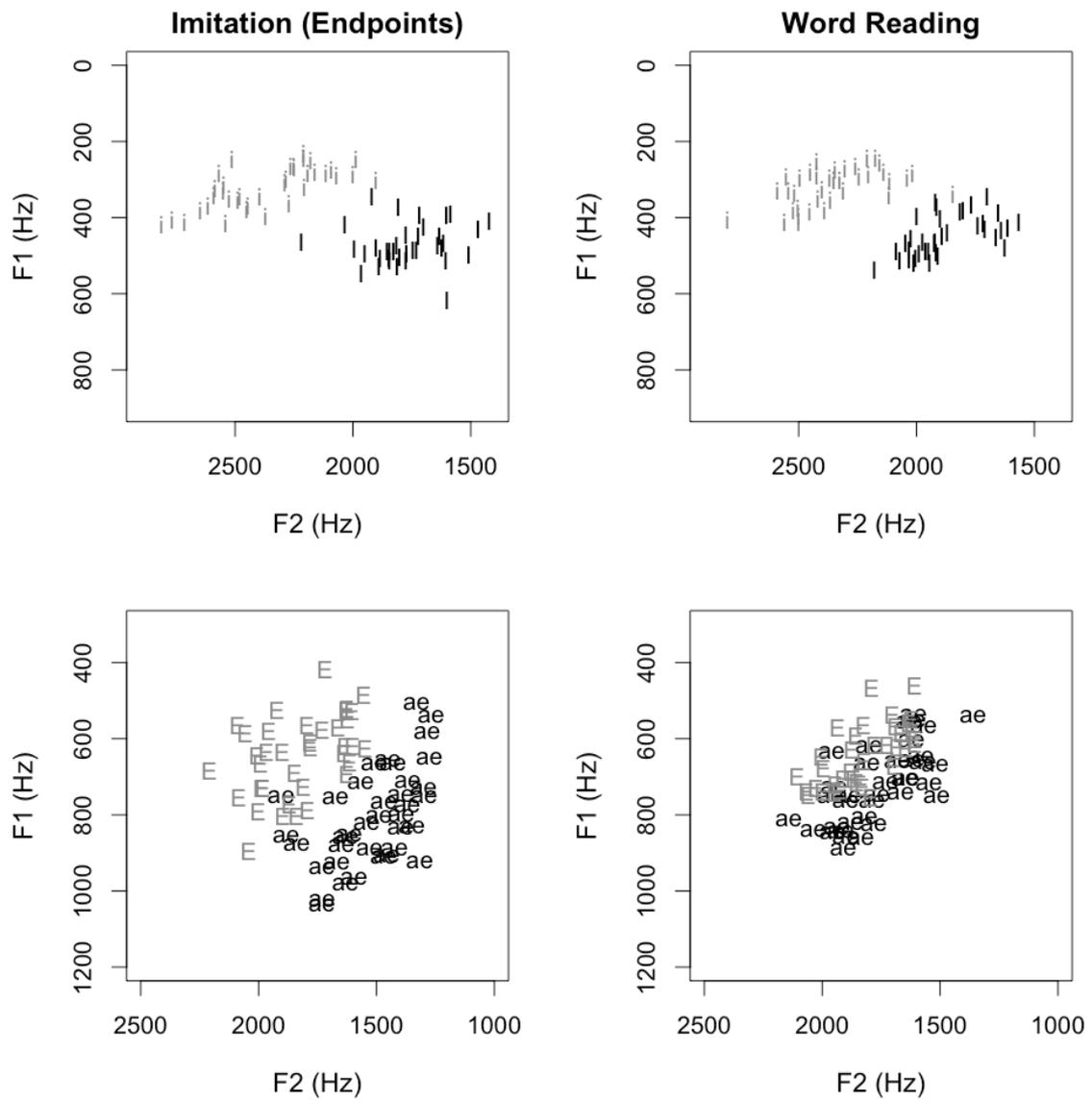


Figure 4

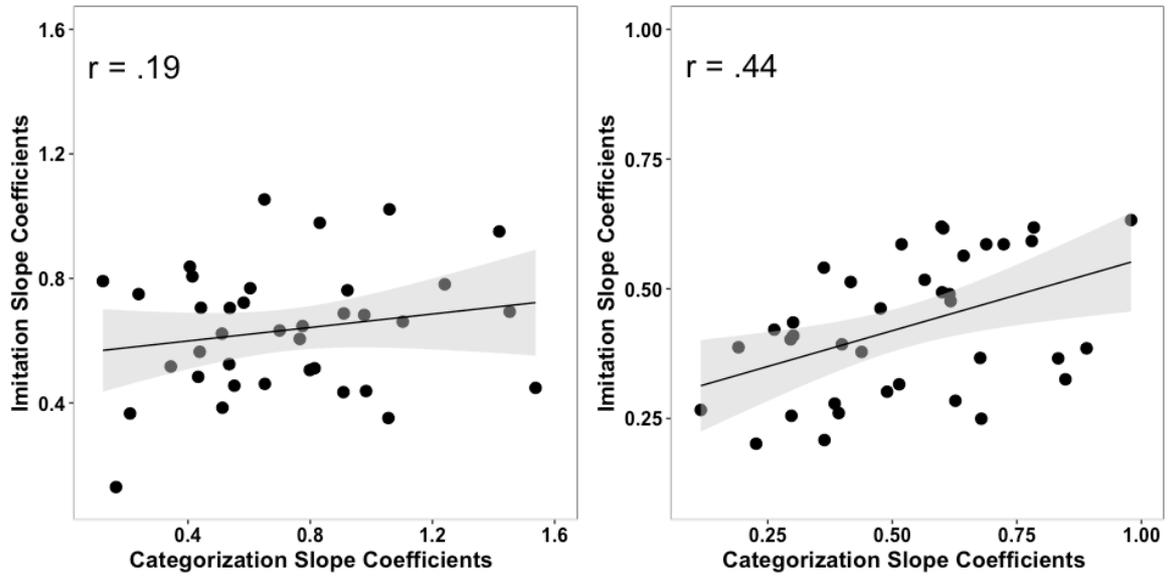


Figure 5

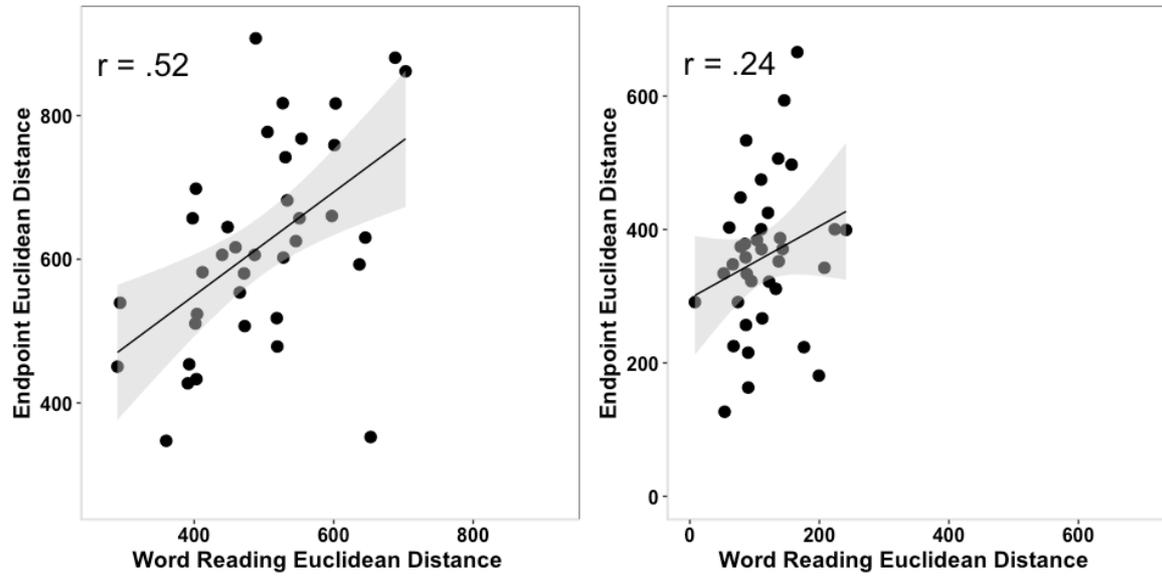


Figure 6

## Appendix A

Table A1

*Words containing /i/, /ɪ/, /ε/ and /æ/ used in the word reading task.*

<i>/i/</i>	<i>/ɪ/</i>	<i>/ε/</i>	<i>/æ/</i>
chief	chill	centre	champion
decent	discount	check	damage
demon	dish	chest	fact
feeling	fiction	death	fashion
field	listen	dentist	hammer
free	listener	desert	hand
frequent	ritual	helmet	happy
genius	scissors	helpful	language
leaf	simple	lecture	national
need	sing	record	Saturday
reader	singer	rest	shadow
secret	thing	smell	smash
speaker	wizard	special	snack