

Full title:

Phonetic alignment in English as a *lingua franca*: Coming together while splitting apart

Authors:

Grant M. Berry

(Department of Spanish, Italian, and Portuguese) Penn State University, PA, USA; (Center for Language Science) Penn State University, PA, USA

Mirjam Ernestus

(Centre for Language Studies) Radboud University Nijmegen, The Netherlands; Max Planck Institute for Psycholinguistics, The Netherlands

Acknowledgements:

Collaboration between authors was facilitated by an NSF Partnerships in International Research and Education grant (#0968369) to Judith F. Kroll, Janet van Hell, Paola E. Dussias, and Ping Li. The second author was supported by a VICI grant (#277-70-010) from the NWO and an ERC Starting Grant (#284108). We thank the Speech Comprehension research group in the Centre for Language Studies at Radboud University and members of the Center for Language Science at Penn State University for their helpful comments and suggestions on earlier versions of this manuscript.

Corresponding author:

Grant M. Berry, 442 Burrowes Building, Department of Spanish, Italian, and Portuguese, Penn State University, University Park, PA, 16802 USA.

Email: grantberry@psu.edu

Abstract

This study investigates the plasticity of phonological boundaries during discourse in a *lingua franca*. We tracked the production of 34 Spanish learners of English conversing with two Dutch confederates in English across two speech styles, focusing on incremental changes in two key English vowel contrasts with differential effects of cross-linguistic influence (/i/-/ɪ/ and /ɛ/-/æ/). Results indicate that Spaniards align with Dutch confederates, quickly merging /ɛ/ and /æ/ and gradually separating their merged /i/-/ɪ/ category, rather than adopting native-like English production. We found greater merger in informal speech overall. We also found an interaction with time for the /i/-/ɪ/ contrast, indicating that the merged /i/ and /ɪ/ categories gradually separate in informal speech; this effect was not found for /ɛ/-/æ/. Finally, proficiency modulates alignment: the most proficient speakers separate /i/-/ɪ/ and merge /ɛ/-/æ/ more than other speakers. We interpret phonetic alignment as a complex, dynamic phenomenon influenced by proficiency in discourse language and speaking style, and whose effects may unfold rapidly or gradually depending on the phonological category investigated.

Keywords

Speech production, phonetic alignment, style-shifting, English as a *lingua franca*, vowel contrasts, conversational speech

Phonetic alignment in English as a *lingua franca*: Coming together while splitting apart

Introduction

Speech in a second language (L2) or additional language provides a valuable opportunity to investigate the plasticity of phonological boundaries (e.g., the /i/-/ɪ/ boundary, which distinguishes ‘sleet’ [sli:t] from ‘slit’ [slɪt]), particularly when that language is learned later in life (e.g., in a classroom setting after the first language has been established). Learner speech is often influenced by pre-existing phonological categories from the first language (L1; cf. Flege et al., 2003; Chang and Mishler, 2012; Burgos et al., 2013; Wilson and Gick, 2014). For example, Spanish native speakers are known to have difficulty making the distinction between /i/-/ɪ/ when learning English, because Spanish does not have this distinction in its phonological inventory. Dutch speakers of English, in contrast, are able to make this distinction in English because it is also present in their native language. Knowledge of a non-native speaker’s L1 phonology thus creates a set of predictable difficulties for speakers from a given language background. In the case where two speakers have distinct native languages and use a third to communicate (i.e., in a *lingua franca*; cf. Trofimovich and Kennedy, 2014, p. 824), the native phonological categories that influence their production are thus also distinct. By focusing specifically on categories that are differentially influenced by the interlocutors’ native languages and measuring changes in those categories in conversational speech, the plasticity of phonological boundaries can be investigated in a naturalistic setting.

The plasticity of phonological boundaries has been well attested in perception. Perceptual boundaries dynamically update in response to various types of phonetic input (e.g., Davis et al., 2005; Kraljic and Samuel, 2007; Vroomen et al., 2007; Maye et al., 2008; Kraljic and Samuel, 2005) and listener expectations about that input (e.g., Hay et al., 2006; Magnuson and Nusbaum, 2007; Kraljic et al., 2008; Hay and Drager, 2010; Babel, 2012; Kraljic and Samuel, 2005; Clopper, 2014). Listeners are able to normalize for foreign or nonstandard accents (e.g., Bradlow and Alexander, 2007; Bradlow and Bent, 2008; Baese-Berk et al., 2013; Clopper, 2014; Witteman et al., 2015; Weber et al., 2014), idiosyncrasies of individual talkers (e.g., Kraljic et al., 2008; Jesse and McQueen, 2011), and the presence of multiple talkers (e.g., Magnuson and Nusbaum, 2007; Kraljic and Samuel, 2005; Kraljic and Samuel, 2007)—among other features. The wealth of research in this area consistently points to a striking conclusion: human beings are exceptionally tuned to process variation in their speech input.

Variation in one’s input may also rapidly translate to variation in one’s production (see, e.g., Kittredge and Dell, 2016). Speakers have been known to adjust vowel quality (Babel, 2010; Delvaux and Soquet, 2007), pitch (Babel and Bulatov, 2011; De Looze et al., 2014), voice-onset time (Shockley et al., 2004; Mitterer and Ernestus, 2008; Nielsen, 2011), speech rate (Pardo, 2001; Sidaras, 2011), and intensity (Pardo, 2001; De Looze et al., 2014) in response to input from an interlocutor. This phenomenon, as it is currently understood, derives from multidisciplinary origins spanning language contact (e.g., Weinreich, 1968; Trudgill, 1986), social psychology (cf. Giles et al., 1991), and cognitive psychology (Chartrand and Bargh, 1999; Pickering and Garrod, 2004). Consequently, terminology regarding the phenomenon has been quite variable. For example, it has been referred to as accommodation (e.g., Trudgill, 1986; Babel and Bulatov, 2011; Babel, 2012; Babel et al., 2013; De Looze et al., 2014), convergence (e.g., Weinreich, 1968; Pardo 2006, Babel, 2010), or alignment (e.g., Pickering and Garrod 2004) in the literature. Here, we adopt ‘alignment’ to refer to the tendency of an interlocutor to adopt production trends from their conversational

partner in discourse (see also Pickering and Garrod, 2004; Costa et al., 2008), because this term has been used to refer to similar phenomena at other levels of linguistic production (e.g., syntax; cf. Branigan et al., 2000).

The bulk of research on phonetic alignment in production has investigated variation within a single phonological category, but individuals are often exposed to systematic variation in production across phonological categories as well. Variation across categories may thus also be context-sensitive and subject to alignment. Simonet (2014) found evidence that Spanish-Catalan bilinguals reduced their production of the Catalan /o/-/ɔ/ contrast when the experimental stimuli also included words from Spanish, which lacks this contrast. Babel et al. (2013) used a shadowing task followed by a perceptual similarity task (AXB) and an acoustic measure of spectral overlap to investigate alignment to a merger in-progress in New Zealand English (the so-called ‘near’ / ‘square’ merger). The researchers found evidence that this sound change-in-progress was partially reversed after participants were exposed to Australian English, which lacks this merger. Similarly, in a study of Korean learners of English, Hwang et al. (2015) found that participants produced more English-like /æ/ and coda /b/ after exposure to a monolingual English confederate. Curiously, participants did not align their production to a Korean-accented confederate who merged /ɛ/ and /æ/ and devoiced coda /b/ to [p] (2015, pp.79-80), even though this speaker and the participants shared the same L1.

When an individual speaks in a *lingua franca*, the production patterns of their interlocutor, which are influenced by that interlocutor’s L1 phonology, may influence the speaker’s own production. In contexts where a native speaker is present as the interlocutor, a non-native speaker may expect the task to center on the accuracy of his/her production, so it is difficult to disentangle phonetic alignment from broad production shifts due to attempting to perform model speech. When a native speaker is absent from the environment, in contrast, it remains an empirical question how bilingual conversational partners may update their production of phonological contrasts while speaking with an interlocutor who has distinct, but nonetheless non-native phonology.¹ If previous findings in phonetic alignment obtain, then we would expect interlocutors to be influenced by one another’s production patterns.

The rate at which alignment occurs in conversational interaction has not yet been investigated in depth, neither for native nor non-native speakers. Alignment may be a continuous process of online updating from perceptual input to speech output, but nearly all research on the topic has followed a baseline—exposure—test approach that assesses alignment at only two time points (cf. Pardo, 2006; Delvaux and Soquet, 2007; Babel and Bulatov, 2011; Nielsen, 2011; Babel et al., 2013; Simonet, 2014; Trofimovich and Kennedy, 2014; Hwang et al., 2015). In fact, nearly no research to our knowledge has investigated phonetic alignment in real time (but see De Looze et al. [2014] for automated analysis of prosodic alignment over time). We hypothesize that exposure should modulate their production of phoneme boundaries throughout the conversation as speakers receive continued input from their interlocutors (cf. Kittredge and Dell, 2016), though this may be affected by the phonological contrast investigated and the overall discourse context.

Conversational speech is known to be affected by contextual factors, such as speech style (e.g., Labov 2001, pp. 106-110, 272; Ernestus et al., 2015; Kouwenhoven et al., 2015; Biber et al., 2006; Biber, 2012). Under a sociolinguistic monitoring account of language processing (Labov et al., 2011) individuals are hypothesized to continuously monitor situational demands in their environment. In formal settings, additional monitoring of the situational demands may strain cognitive resources such as working memory and attention (cognitive processing load) further than

in an informal setting. For example, Mattys and Wiget (2011) found that the tendency for lexical information to bias perception of phonological contrasts (the “Ganong” effect; Ganong, 1980) increased when participants were given a visual search task alongside the perception task, thus increasing cognitive processing load. Furthermore, the influence of lexical information on perception did not decrease over time in the high-load condition. When attention is divided during speech processing, individuals are less able to dedicate cognitive resources to processing fine phonetic detail, which in turn diminishes the influence of phonetic information present in the task on performance. Thus, when cognitive resources are divided during language processing, individuals may rely more on generalized, abstract (e.g., lexical) representations than fine phonetic detail (cf. Mattys and Wiget, 2011; Mattys et al., 2014; Mattys and Palmer, 2015).

In production, discourse context exerts considerable influence on the use of linguistic variables (e.g., Levelt, 1983; Labov, 2001; Labov et al., 2011). For example, Thompson and Brown (2012) asked a Spanish L2 learner of English to participate in five speech elicitation tasks ranging from informal to highly formal, in the style of a sociolinguistic interview. These tasks included personal narratives and free speech, an English conversation with the participant’s husband, reading a passage aloud, reading word lists, and pronunciation of minimal pairs. The researchers focused on the Spaniard’s production of the English lax /ɪ/ (as in ‘bit’), since this vowel is notably difficult for Spanish speakers of English, who only have /i/ in their phonological inventory (see also Flege, 1991; Konaurova and Francis, 2008; Casillas, 2015). The researchers found that the number of words containing /ɪ/ where this vowel was produced incorrectly (e.g., as /i/) increased significantly as a function of task formality, ranging from 4% in the personal narrative to 58% in the production of minimal pairs. These findings suggest that speech in a formal register may involve stronger influence from the more familiar L1 phonology. Regarding phonetic alignment, additional monitoring and decrease in available cognitive resources (e.g., Mattys and Wiget, 2011; Thompson and Brown, 2012) may result in a diminished ability to align to the interlocutor.

Cognitive processing load in a *lingua franca* may additionally be modulated by proficiency in that language. More proficient speakers are able to more fluidly process linguistic input in their non-native language, which may decrease cognitive processing load (see Garcia Lecumberri et al., 2010, for a detailed review of non-native perception in adverse conditions). For example, Serafini and Sanz (2015) tested a group of L1 English learners of Spanish using online grammatical tasks, and the researchers found that the influence of phonological working memory was inversely correlated to L2 proficiency. In online priming tasks using event-related potentials (ERPs), Elston-Güttler et al. (2005) found evidence that L1 activation gradually decreased as proficiency increased (Elston-Güttler et al., 2005, p. 1605; see also Moreno et al., 2008, for a review of EEG studies on bilingual phonology and the role of proficiency in modulating ERP responses). Additionally, White et al. (2017) recorded EEG responses as monolingual English speakers and French-native late bilinguals of English performed phoneme discrimination, phoneme categorization, and pseudoword reading tasks. The researchers found that high proficiency bilinguals showed similar ERP signatures to native speakers, demonstrating sensitivity to L2 phonological contrasts, but lower proficiency bilinguals did not show sensitivity to the L2 contrast. Broadly, these results suggest that language proficiency impacts language processing on a neurocognitive level, such that higher proficiency leads to more native-like processing.

If higher proficiency relates to more fluid processing and less monitoring, then this would free up cognitive resources to attend to fine phonetic detail in the discourse context. Less proficient speakers, who have more difficulty processing speech input both within and across phonological

categories (see also Bohn and Flege, 1990; Flege et al., 1997; Archila-Suerte et al., 2012), may thus also have reduced capacity to focus on idiosyncrasies of their interlocutor's speech, particularly when that interlocutor's phonology possesses a phonemic contrast with which they are unfamiliar. For this reason, more proficient speakers may demonstrate greater and/or faster alignment than less proficient speakers (see also Kim et al., 2011).

In the current study, we track Spanish native speakers' production of two key vowel contrasts in Standard Southern British English as they converse with two Dutch interlocutors. The first contrast, /i/-/ɪ/ (e.g., 'beet' vs. 'bit'), is difficult for native speakers of Spanish (Flege, 1991; Flege et al., 1997; Casillas, 2015; Barrios et al., 2016), which does not have a tense-lax vowel distinction (e.g., Martínez-Celdrán et al., 2003), but not for native Dutch speakers, who have this contrast in their L1 inventory (e.g., Booij, 1995, p. 5). The second vowel contrast, /ɛ/-/æ/ (e.g., 'pen' vs. 'pan'), was chosen for the opposite reason: this contrast is reliably produced by Spanish speakers of English, but difficult for Dutch speakers of English. Spanish learners of English produce English /ɛ/ (often diphthongized to [eɪ]) phonetically closer to the space of monolingual English /ɛ/ (e.g., Bradlow, 1995, p. 1919; Giacomino, 2012). Additionally, their English /æ/ is slightly backed from monolingual English targets (Giacomino, 2012), moving it toward the Spanish /a/ and farther away from the space of /ɛ/. Though the production of L1 Spanish speakers of English may remain accented, their English vowel categories are nonetheless reliably distinguished from one another in production and perception (cf. Flege et al., 1997; Escudero and Chládková, 2010; Archila-Suerte et al., 2012), which is the feature of interest in the current study. Converseley, Dutch only has /ɛ/ in the lower front quadrant of the vowel space (e.g., Gussenhoven, 1992; Booij 1995, pp. 4-5). In Dutch English, /ɛ/ and /æ/ consistently merge in production (e.g., Adank et al., 2004b; Wang and Van Heuven, 2006; Adank et al., 2007), even though Dutch speakers can discriminate these phones (cf. Simon and D'Hulster, 2012; Escudero et al., 2012).

When native speakers of Spanish, who are the focus of the current study, converse with Dutch interlocutors in English as a *lingua franca* (ELF), they are exposed to two variables not present in their L2 production: an English-like vowel contrast they do not habitually produce (/i/-/ɪ/) and a merger of two distinct English vowels that they typically distinguish (/ɛ/-/æ/).² We envision two scenarios. Alignment to the Dutch interlocutor would necessitate adoption of both of these variables, though partial alignment may also occur solely via merger of the /ɛ/-/æ/ contrast (which the Spanish speakers separate in their L1), here measured by a multivariate metric that approximates the shared area of two vowels spanning the first and second formants (the Pillai score; cf. Hay et al., 2006; Babel et al., 2013; Amengual and Chamorro, 2015). Conversely, orientation to more native-like English speech would result in adoption of solely the /i/-/ɪ/ distinction. Since no native speakers are present in the discourse environment, there is no obvious bias toward native-like speech in these contexts. We would therefore expect the first scenario: Spanish learners align to the speech patterns of their Dutch interlocutors by allowing their /ɛ/-/æ/ categories to come together while their merged /i/-/ɪ/ category gradually splits apart.

The reported study uses a new conversational corpus in English as a *lingua franca* (the Nijmegen Corpus of Spanish English; Kouwenhoven et al., 2015) to explore phonetic alignment as a function of both time and speech style, focusing on dynamic variation in the realization of phonological contrasts. Based on the previous findings in bilingual language processing and monolingual phonetic alignment aforementioned, this study formulates and tests the following four hypotheses:

- (H1) Spaniards align to their Dutch interlocutors by separating /i/-/ɪ/ and merging /ɛ/-/æ/;
- (H2) Continued exposure results in more alignment, reflected as a main effect of time;
- (H3) There is less alignment in formal than informal speech; and
- (H4) More proficient speakers align to their interlocutor more than less proficient speakers, reflected by a main effect of proficiency and/or an interaction between proficiency and time.

Methodology

The Corpus: The Nijmegen Corpus of Spanish English

The Nijmegen Corpus of Spanish English (NCSE; Kouwenhoven et al., 2015) is a novel corpus of bilingual speech collected in Madrid in 2014. The NCSE consists of recorded conversations between 34 Spanish learners of English (17 females and 17 males, aged 19 to 25 years) in informal and formal conversation with two Dutch confederates (one male, aged 23, and one female, aged 24). Most of the Spanish-English participants were near the end of their undergraduate degree program ($\bar{x} = 3.66$, $s = 1.41$ years of study) and nearly all ($n = 30$) were students of engineering. Importantly, the Spanish participants were not recruited based on their English fluency. On average, the participants reported studying English for 12 years ($\bar{x} = 12.07$, $s = 5.16$), which is consistent with the obligatory study of foreign languages in secondary school in Spain. In contrast to the Spanish participants, the confederates were chosen to have intermediate proficiency in English and specifically not to sound like native speakers. Because the confederates did not sound native, there is no *a priori* reason to expect the Spanish speakers should treat the confederates' English as model speech.

In the informal recordings, sex-matched pairs of Spanish L1 and Dutch L1 speakers engaged in unrestricted conversation in English as a *lingua franca*. The informal recordings were initiated under the guise that both the Dutch confederate and the Spaniard were participants in an experiment who were waiting for the experimental materials to be brought in to the testing room; this created a speech situation where both the participant and the confederate were effectively peers. The confederates were instructed to allow the Spaniards to talk most of the time to elicit more speech from the participants than from the confederates, which they did (see Appendix 1). When the impromptu conversation naturally came to an end, the experimenter brought in a stack of cards with names of famous people and places to facilitate further informal discourse. The confederate and participant then played a guessing game, taking turns selecting a card and describing the item to one another until it was successfully named or the conversational partner passed and moved to the next item.

After the informal recordings, the same Spanish learners of English were given a written description of the formal recordings to follow. They were told that these interviews were being conducted for a Master's thesis in journalism on the recent economic crisis in Spain and Europe. Participants were informed that their responses would be compared to those of politicians and other influential individuals. To maximize the effect of formality, the participant then changed into formal attire. A video camera was placed on a tripod and pointed directly at the participant, and the participant was interviewed by a confederate of the opposite sex.³ Evidence that the formality manipulation was successful was found in Kouwenhoven et al. (2015), where speech style modulated the amount of laughter, overlapping speech, information content, and use of Spanish (L2) words in the conversation, among other features.

Once the recordings were collected, an ESOL teacher trained in the Common European Framework of Reference for Languages (CEFR) rated the proficiency levels of the Spanish-English participants and the Dutch confederates using that scale; a +/- system was used to rate speakers that fell between standard CEFR categories (see Council of Europe, 2001, p. 24). The Dutch confederates spoke at an intermediate to advanced intermediate level (between B1 and C2), while the Spanish speakers ranged from a basic elementary (A1) to an upper intermediate level (B2). Eight different proficiency levels were represented by the Spanish participants: A1 (two speakers), A2 (eight speakers), A2+ (two speakers), B1- (three speakers), B1 (11 speakers), B1+ (five speakers), B2- (two speakers), and B2 (one speaker).

Data collection: Forced Alignment and Vowel Extraction (FAVE)

All tokens of the vowels of interest were isolated and extracted using FAVE (Rosenfelder et al., 2014)⁴. FAVE utilizes Hidden Markov Modeling in combination with pronunciation models based on American English to demarcate an audio signal into words and phones based on a word-for-word transcription of the speech in that signal.⁵ FAVE then extracts F1 and F2 values from the audio files by vowel using the demarcated signal, and it normalizes those values using a log-mean method; this is a vowel-extrinsic measure that reduces anatomical and physiological variation (cf. Adank et al., 2004a; Rosenfelder et al., 2014). The computational process for isolating vowel formants trades human error for mechanical error, but the mechanisms in the latter are consistently and more objectively applied than in the former. Additionally, the software permits the collection of much more data than would otherwise be possible in the same amount of time. The larger dataset permits a more nuanced analysis of interesting linguistic phenomena and offers greater statistical power.

To eliminate extreme outliers or errors resulting from the computational method of vowel extraction, all tokens greater than 2.5 standard deviations about the mean in F1 and F2 were excluded; this was done on the Dutch and Spanish data in aggregate. For the Spaniards, this process resulted in a loss of 3.6% (2,401 tokens) overall. By vowel, the process resulted in a loss of 2.7% of the data for /i/, 4.7% for /ɪ/, 2.8% for /ɛ/, and 3.5% for /æ/. We retained 8,868 tokens of /i/, 18,636 tokens of /ɪ/, 21,057 tokens of /ɛ/, and 17,405 tokens of /æ/. For the Dutch, this resulted in a loss of 3.9% (1,542 tokens) overall, which constituted a 2% loss for /i/, a 5.2% loss for /ɪ/, a 2.5% loss for /ɛ/, and a 4.5% loss for /æ/. We retained 6,054 tokens of /i/, 10,672 tokens of /ɪ/, 11,268 tokens of /ɛ/, and 11,646 tokens of /æ/ (see also Appendix 2).

Because proficiency levels were assessed post hoc and therefore could not be balanced, we re-categorized proficiency into three groups: Lower (A1, A2, A2+; 12 speakers); Middle (B1-, B1; 14 speakers), and Higher (B1+, B2-, B2; 8 speakers).

The data as time series

To evaluate whether alignment in ELF speech is incrementally modulated by exposure to an interlocutor, we utilized statistical models that permit the inclusion of fixed effects as well as their interactions with time.

Time was rescaled from 0 to 100 for each recording, which was then subdivided into five equally spaced time bins.⁶ Informal conversations were longer than formal conversations ($t(33) = 65.67$, $p < 0.001$; $\bar{x}_{\text{Formal}} = 27.51$ min and $\bar{x}_{\text{Informal}} = 50.23$ min), and token counts per time bin could therefore not be controlled. To correct for this, time bins with fewer than 30 tokens were excluded from the analysis.⁷

Measuring degree of category separation: The Pillai score

We quantified the spectral overlap between the vowels in each contrast with Pillai scores (Pillai 1954, 1955; Olson, 1976; Hay et al., 2006); these were calculated for each speaker, conversation, time bin, and vowel contrast using an R script written by the first author (R Core Team, 2016; available in online supplemental materials).⁸ To our knowledge, Hay et al. (2006) were the first to use Pillai scores to quantify degree of vowel merger. The metric has been used in this way in subsequent work, such as for measuring alignment to English mergers-in-progress (Babel et al., 2013) or exploring the degree of cross-linguistic competition in bilingual vowel production (Amengual and Chamorro, 2015). In a recent meta-analysis of metrics for vowel merger, Nycz and Hall-Lew (2013) confirmed that Pillai scores adequately index spectral overlap.

The Pillai score, also known as Pillai's criterion or the Pillai-Bartlett trace, is one of the output components of MANOVA that measures how well the independent variables explain the dependent variable. The Pillai score is based on the product of the variance matrices for the model fit and model error, and can be calculated as the summation of each of the p eigenvalues in this product divided by its value plus one (see Equation (1)).

$$\text{Pillai Score: } \sum_{i=1}^p \frac{\lambda_i}{1+\lambda_i} \quad (1)$$

For the Pillai score to index degree of merger, ordered pairs of F1 and F2 values are used as the outcome variable in the MANOVA. These ordered pairs, which indicate the tongue position of the vowel tokens, are predicted by the English vowel category label (e.g., /i/ vs. /ɪ/), which is added as the independent variable. When MANOVAs are conducted in this way, the Pillai score quantifies the degree of overlap of the two vowels in F1 x F2 space. In this case, where Vowel Category is the only independent variable, $p = 1$, and the Pillai score ranges from zero to one.⁹ Scores close to zero indicate that the vowel category label accounts for very little variance in F1 and F2 in the data, and thus imply that the categories are strongly merged. Conversely, scores close to one indicate that the vowel category label strongly explains the observed variance in F1 and F2, and hence provide evidence that the two categories do not overlap. Importantly, the Pillai score is a conservative measure relative to other common MANOVA outputs, and it is quite robust to imbalanced data and violations of MANOVA assumptions (Olson, 1976).

Selecting the dependent variable: Spanish Pillai, Dutch Pillai, or their difference?

We refer to alignment as the process of updating one's speech patterns in response to one's exposure. Because both interlocutors receive continued exposure from their conversation partner, presumably both align their speech to their input in some respect. In the current study, we focused on changes in the Spaniards' phonological category production, because they were the group of interest in designing the NCSE corpus. The number of tokens from the confederates, who were not the primary focus of the conversations, was not sufficient to permit the calculation of Pillai scores for several time bins at our predetermined threshold. Regrettably, this made conducting a difference analysis including a three-way interaction among Style, Proficiency Group of the Spaniard, and Normalized Time untenable. We did, however, verify that changes in the Dutch interlocutors' Pillai scores and their normalized F1 and F2 values were minimal across time and

speech style before conducting the analysis, both through visual examination and linear mixed effects modeling of Dutch Pillai scores (see Appendix 3).¹⁰

For the Spanish participants, we studied the Pillai difference at each time bin, for each speaker and speech style, using this as an index of alignment. We then interpreted the results to present a robust description of the changes in phonological category production that occurred.

Statistical analysis: Linear mixed effects models

To predict the dependent variable (Spanish Pillai score), we fit linear mixed effects models to the data using the `lme4` package in R (Bates et al., 2015a; R Core Team, 2016).¹¹ We began with models containing the theoretically warranted predictors (Proficiency, Speech Style, Time, and all interactions), a random intercept by speaker, and all random slopes justified by the model (Barr et al., 2013). According to Barr et al., the appropriate random slopes would be any factors expected to vary within-speaker. In this case, we might reasonably expect speakers to be differentially sensitive to speech style (i.e., speakers may respond differently to discourse pressures in each setting) or Time Bin (speakers may align at different rates from one another), and these effects may interact with one another (i.e., a speaker who is strongly affected by speech style might also be strongly affected by time). Proficiency Group was coded with orthogonal polynomial contrasts to permit an evaluation of linear and quadratic trends in proficiency on degree of category separation. Additionally, Speech Style was orthogonally sum-coded to center the estimate on the grand mean across styles. We then systematically refined the model structure, removing higher-order effects that were not significant and checking model fits with likelihood ratio tests.

To simplify the random effects structure and eliminate redundancies in the random slopes, Bates et al. (2015b) recommend that a Principal Component Analysis (PCA) be conducted on the random effects in the model. We used the `rePCA()` function in the `RePsychLing` package (Baayen et al., 2015) to accomplish this. Conservatively, we selected the number of principal components that accounted for at least 99% of the variance in the random effects structure. We examined factor loadings of each random slope and systematically removed any that correlated to the selected principal components with an absolute value less than 0.4, with the stipulation that all main effects be present for any interaction in the random effects structure. We then checked this model for convergence and standard diagnostics of model fit. Regarding the interpretation of the mixed model, we assume that the *t*-statistic resulting from the linear mixed effects models has approximately a standard normal distribution; thus, we accept all *t*-values with magnitude greater than 2.0 as significant (see also Baayen et al., 2008, p. 398, footnote 1).

Rather than conduct a single linear mixed effects model on the full dataset and include Vowel Contrast in the fixed effects structure, we conduct separate models for each contrast. This was done because the two contrasts in English, which occupy distinct regions in an asymmetrical vowel space, may have distinct degrees of spectral overlap to begin with. Thus, a small increase in Pillai score for vowels with higher degrees of overlap may be interpreted differently than the same increase for vowels that typically show a lower degree of overlap. Although very little work has attempted to characterize spectral overlap among native speakers' production of Standard Southern British English vowels, research on West Midlands English has shown that /i/-/ɪ/ has more overlap than /ɛ/-/æ/ in this dialect group (Sharifzadeh et al., 2012, pp. e51-e52). For American English in Michigan, Hillenbrand et al.'s (1995) study shows the opposite: participants showed greater overlap for /ɛ/-/æ/ than /i/-/ɪ/. Regardless of the directionality of the difference, we assume that the

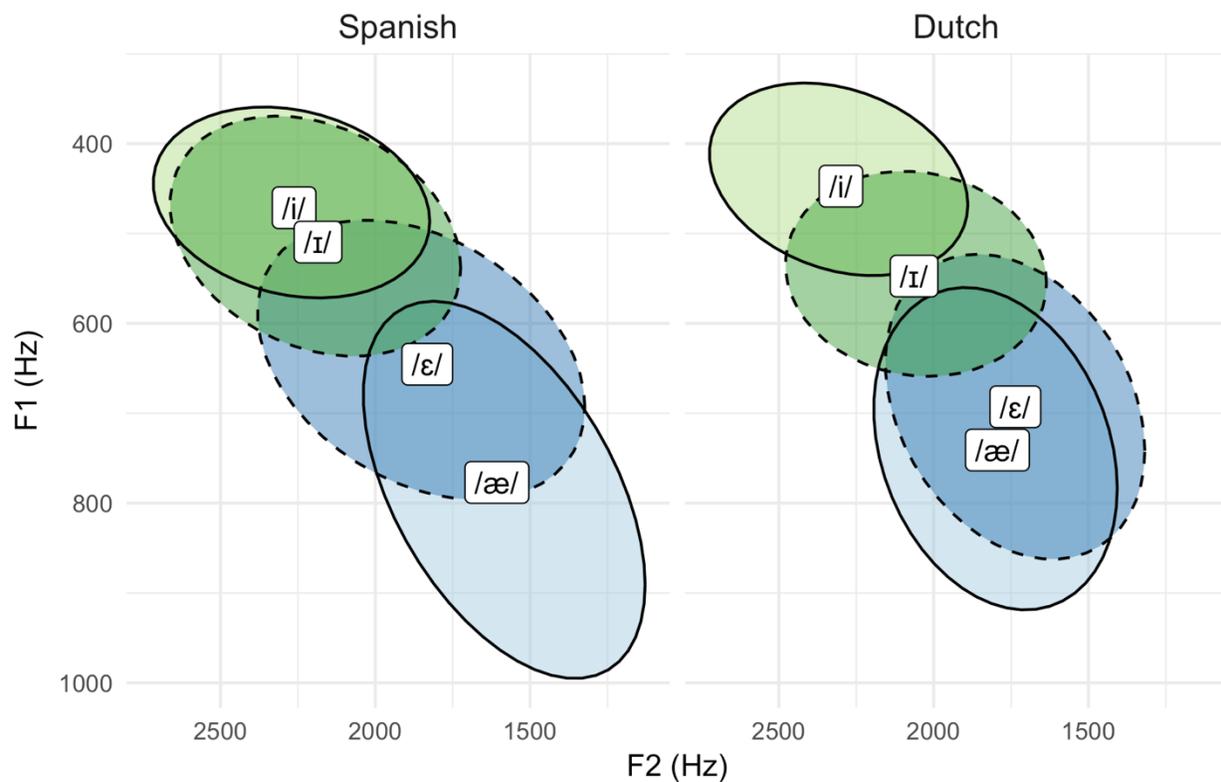
two target vowel contrasts have distinct degrees of separation which would not be adequately captured in a composite model that combines both vowel contrasts.

Results

Confirmation of L1 transfer

We begin by plotting the vowel pairs by L1 to visually confirm that the Dutch confederates' and Spanish participants' English production in aggregate demonstrates transfer from their L1 phonology. Figure 1 plots spectral overlap for all vowels extracted from the conversations pertaining to the phonological contrasts of interest for the Spaniards and the Dutch confederates, respectively; these represent vowel productions over the entire course of the experiment. The two L1 Dutch confederates merge /æ/ and /ɛ/ while maintaining the contrast between /i/ and /ɪ/, and the L1 Spanish participants as a group do the opposite— as expected. That the Dutch confederates merged the /æ/ and /ɛ/ contrasts provides additional evidence that they did not sound like native English speakers. Since the Spaniards do make this contrast, this fact should have been easily noticeable to the participants.

Figure 1. *Spectral overlap in target vowel pairs for aggregate data by native language (Vowel labels mark centroids; line types differentiate two members of a target contrast; ellipses are 95% confidence regions about the mean in logmean-normalized F1x F2 space)*



Factors influencing Spanish category separation

Before interpreting the statistical models of the Spaniards' raw Pillai scores, we first checked standard model diagnostics. Quantile-quantile plots confirmed that the residuals were normally distributed, and the autocorrelated residuals were low (-0.1 to 0.1 for most lag values). The former is an assumption of linear regression, while the latter indicates that the model did not overestimate its parameters, despite the non-independence of the observations. With these established, we reviewed the model outputs. The fixed effects structure for the /i/-/ɪ/ contrast included main effects for Speech Style, Normalized Time Bin, and Proficiency, as well as an interaction between Speech Style and Time (see Table 1). The random effects structure included a random intercept by speaker and a random slope by Speech Style.

Table 1. Mixed model results for Spaniards' production of /i/-/ɪ/ contrast (N = 323).

Predictor	β-Estimate	Std. Error	t-value
Intercept	0.144	0.014	10.228
Time Bin	-0.002	0.003	-0.565
Informal Style	-0.072	0.020	-3.645
Proficiency (Linear)	0.045	0.020	2.276
Proficiency (Quadratic)	0.030	0.018	1.691
Informal Style : Time Bin	0.017	0.006	2.809

The main effect of Speech Style indicates that /i/ and /ɪ/ were more merged in the informal recordings than in the formal recordings (see Figure 2). There was also a linear effect of proficiency, indicating that separation of /i/ and /ɪ/ increased with proficiency (see Figure 3). While there was no main effect of time bin, there was an interaction between Speech Style and Time Bin. Pillai scores were lower in informal speech than in formal speech, but they increased over time in informal conversation, suggesting that gradual exposure to the Dutch interlocutor in an informal context increased the degree of separation for this contrast. (see Figure 4).

Figure 2. Distribution of Pillai scores by speech style and vowel contrast.

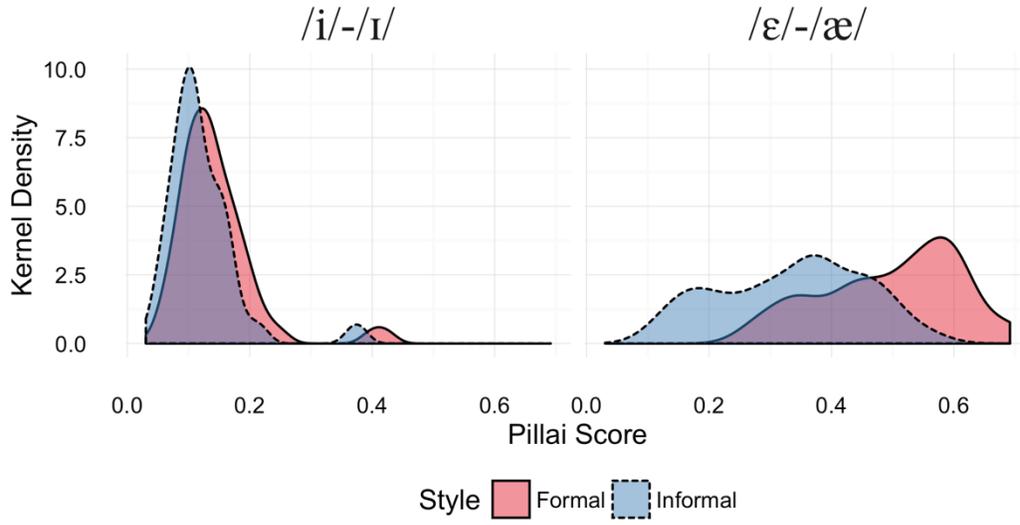


Figure 3. Distribution of Pillai scores by speech style, vowel contrast, and proficiency group

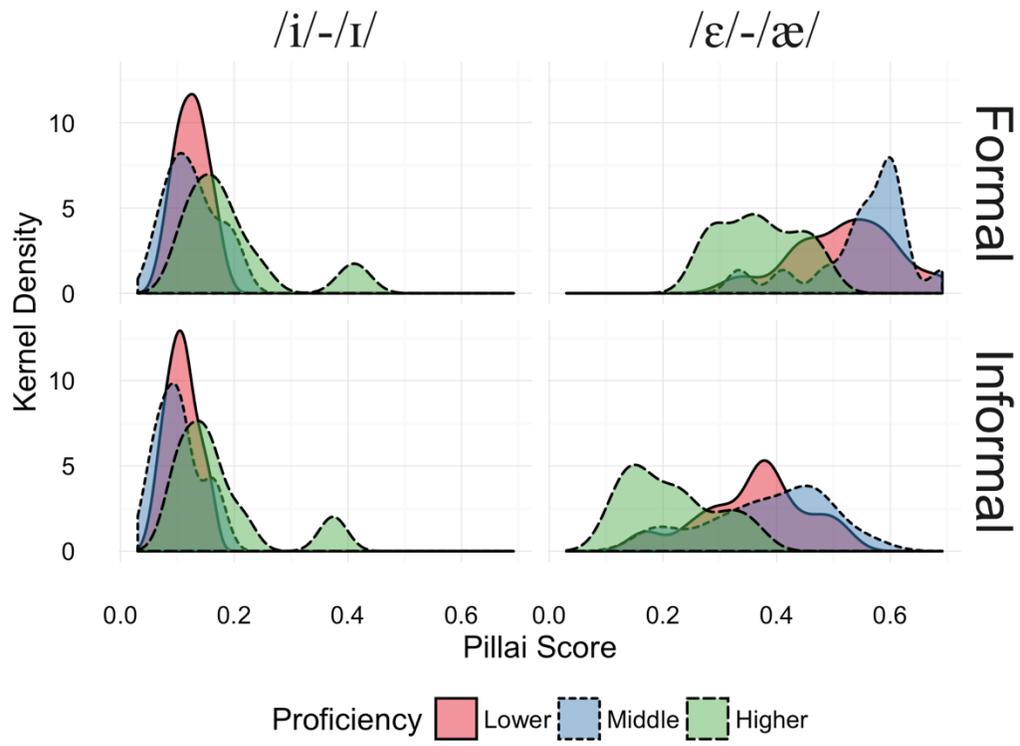
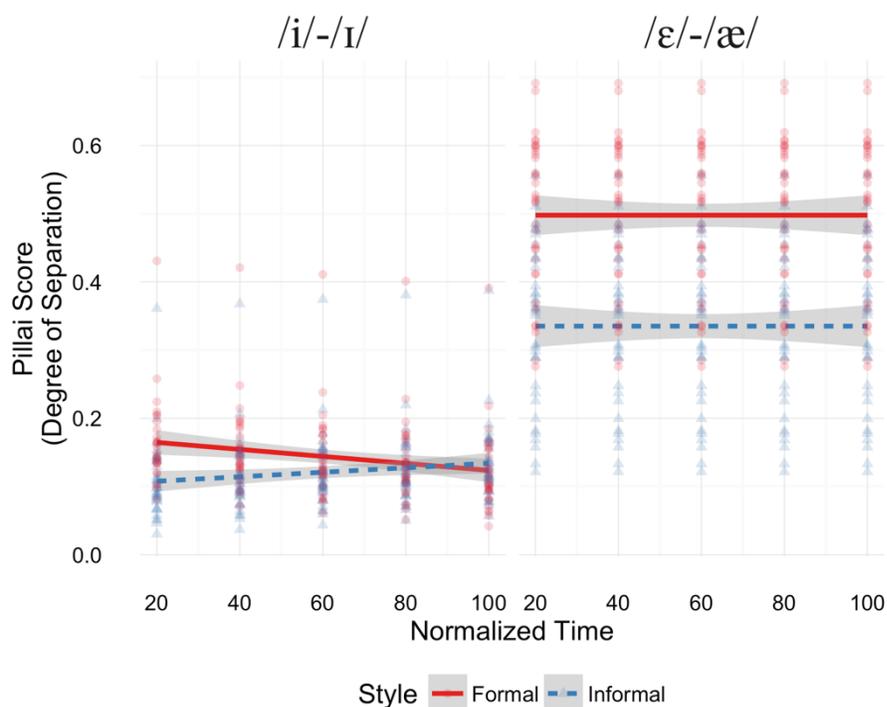


Figure 4. Pillai scores by speech style, time bin, condition, and contrast.

The model based on the $/\varepsilon/-/\text{æ}/$ contrast consisted only of main effects for Speech Style and Proficiency; there was no effect of time (see Table 2). Nonetheless, the random effects structure was identical to the other model (by-speaker random intercept with a random slope for Speech Style). There was no main effect of Time Bin, nor was there a significant interaction between Time and Speech Style (see Figure 4). We did, however, find a main effect of Speech Style, such that $/\varepsilon/$ and $/\text{æ}/$ were more significantly merged in informal speech (see Figure 2).

The model for the $/\varepsilon/-/\text{æ}/$ contrast also revealed linear and quadratic effects for proficiency, such that $/\varepsilon/$ and $/\text{æ}/$ merged significantly more as proficiency level increased (see Figure 3). The finding of both a linear and quadratic relationship of proficiency for the $/\varepsilon/-/\text{æ}/$ contrast indicates that differences among proficiency levels are larger for this contrast than for the $/i/-/ɪ/$ contrast, where only a linear relationship was found. The directionality of the effect is unexpected, because for each contrast in English, higher proficiency would be expected to result in more native-like pronunciation and thus higher Pillai scores (i.e., more category separation). On the contrary, the most proficient learners merged the $/\varepsilon/-/\text{æ}/$ contrast more than speakers of lower proficiency groups (see Figure 3), in line with Dutch English.

Table 2. Mixed model results for Spaniards' production of $/\varepsilon/-/\text{æ}/$ contrast ($N = 340$).

Predictor	β -Estimate	Std. Error	t -value
Intercept	0.399	0.017	23.152
Informal Style	-0.163	0.014	-11.392
Proficiency (Linear)	-0.106	0.031	-3.387
Proficiency (Quadratic)	-0.086	0.028	-3.096

The model findings suggest that production of the two contrasts was conditioned by distinct factors. The two models find greater merger in informal speech as opposed to formal speech, resulting in category separation more like Dutch-accented English for the /æ/-/ɛ/ contrast, but not the /i/-/ɪ/ contrast (which would be split). The interaction with time implies that the /i/-/ɪ/ categories gradually separate in informal speech. This gradual effect was not found for the /ɛ/-/æ/ contrast. Additionally, higher proficiency correlated to contrast production that better matched Dutch-accented English than Spanish-accented English (lower Pillai scores for /ɛ/-/æ/, higher Pillai scores for /i/-/ɪ/), particularly for the /ɛ/ and /æ/ categories.

To summarize, the current study finds evidence that non-native speakers communicating in a *lingua franca* update their production of phonological contrasts in conversational speech. Spanish speakers of English as a *lingua franca* gradually separated their /i/-/ɪ/ category with exposure to their Dutch confederates who make this distinction, but Spaniards only separate this category in informal speech. The Spanish participants also show increased merger for a distinction in English that they can make (/ɛ/-/æ/) during informal conversation with their Dutch interlocutors, who have difficulty with this contrast. We found a robust effect of proficiency level in the merger of these categories, indicating that these effects were stronger for speakers of higher English proficiencies.

Discussion

This study uses the context of English as a *lingua franca* to investigate how bilingual speech production might dynamically update in response to an interlocutor from a distinct language background. Using a new corpus (NCSE; Kouwenhoven et al., 2015), we tracked the production of two key phonological contrasts as 34 Spanish learners of English conversed with Dutch confederates in the context of English as a *lingua franca*. One contrast, /i/-/ɪ/, was selected for its reported difficulty for Spanish speakers of English, but not for the Dutch English confederates. The other, /ɛ/-/æ/, was chosen because it is reliably produced by Spaniards and is reportedly merged in Dutch speakers' English. We measured logmean-normalized F1 and F2 frequencies of the target vowels throughout each conversation. We then rescaled time from 0 to 100 and subdivided the recordings into five equally spaced time bins. We utilized the Pillai score as an index of the degree of separation in F1 x F2 space for the vowels in each contrast, and we conducted statistical analyses to investigate factors modulating those contrasts for the Spanish learners of English.

Our first hypothesis was that Spaniards would align to the speech of their Dutch interlocutors as opposed to the English spoken by native speakers. The former would have required merger of the /ɛ/-/æ/ distinction and development of category boundary between /i/ and /ɪ/, while the latter would have resulted in a four-way distinction among the phones under investigation. In support of our hypothesis, we found that the Spanish participants suspended a category distinction they can make (/ɛ/-/æ/). They also aligned to their Dutch interlocutors in their production of /i/ and /ɪ/ by separating their merged category incrementally over time, though this only occurred in informal speech. These findings suggest that speakers' production of phonological contrasts in English as a *lingua franca* does not gravitate toward abstract, native-like speech norms, but rather to the production patterns that are present in the discourse environment. Thus, we find support for our first hypothesis, although the degree and rate of alignment to the Dutch interlocutor varied by contrast.

Our second hypothesis was that the effect of time would be significant for both contrasts, but we found that time modulated the production of only the /i/-/ɪ/ contrast, and only in informal speech.

This suggests that time-sensitive changes due to alignment only occurred for the more difficult /i-/ɪ/ contrast in informal speech, where cognitive processing load due to contextual monitoring may be diminished.

We also hypothesized that there would be more alignment in informal speech overall, chiefly because lower monitoring in this style would result in lower cognitive processing load than in formal speech. In line with this hypothesis, the /ɛ/ and /æ/ categories were significantly more merged in informal speech, as were the /i/ and /ɪ/ categories.

Informal speech has been found to consist of more reduced forms in general (e.g., Ernestus et al., 2015), which may have contributed to the greater merger observed for each contrast ('hypoarticulation'; cf. Lindblom, 1990). Hence, had we only focused on the /ɛ-/æ/ contrast, which involves collapsing a pre-existing distinction, gestural reduction may have been an alternative explanation for the observed increase in merger. However, the results from the /i-/ɪ/ contrast suggest that gestural reduction is not the sole process underlying alignment in informal speech. Although /i-/ɪ/ was more merged toward the beginning of the informal recordings than in the formal recordings, the interaction with time reveals that merger decreased throughout the informal conversations. Since this effect was not present in formal speech, the interaction provides evidence that phonetic alignment occurs in informal speech modulo the presence of reduced forms. If gestural reduction were a driving force in the observed merger, we would have also expected a negative effect of time on Pillai score in both speech styles (i.e., less differentiation of the two contrasts over time). This should particularly be true in the informal data, where gestural reduction is typically higher. However, we do not find such an effect. Thus, a simple reduction account may be insufficient to explain the observed patterns. Given that the Dutch confederates show no style shifting to the /æ-/ɛ/ contrast, but the Spaniards show a robust effect in the direction of Dutch production in informal speech, we believe this strongly suggests that the speakers showed alignment for this contrast—even though we do not see an effect of time.

Our final hypothesis was that proficiency would influence the degree of alignment. More proficient speakers separated the /i-/ɪ/ contrast to a higher degree than less proficient speakers. The absence of interactions between Proficiency and Time Bin and between Proficiency and Style suggests that the time-sensitive consequence of alignment observed in informal speech occurred at an equal rate for speakers of each proficiency level. This observation in isolation would suggest that alignment is not modulated by proficiency. However, results from the /ɛ-/æ/ contrast suggest the opposite may be true. Speakers from higher proficiency levels more strongly merged their /ɛ/ and /æ/ categories, even though these are largely distinguished in Spanish speakers' L2 English. Additionally, the difference analysis showed that proficiency modulated the difference between interlocutors in category separation for /ɛ/ and /æ/ in an approximately quadratic relationship.

One possible explanation for this finding is that more proficient speakers of a second language, who process speech more fluidly, may have had more cognitive resources available to process information present in the acoustic signal. This would allow speakers of higher proficiencies to attend to the differences between the Dutch interlocutor's production and their own. If differences in the Dutch interlocutors' usage of linguistic variables is tracked in real-time, as a sociolinguistic monitoring account would predict (e.g., Labov et al., 2011), then an expected result is that those production patterns would be integrated into the speaker's mental representations and influence their subsequent production. Higher proficiency speakers would then be expected to suspend the /ɛ-/æ/ contrast, which they are able to make, more so than speakers of lower proficiencies (see also Kim et al., 2011).

Alternatively, higher proficiency speakers may have more native-like phonetic targets in their English production. Since English /ɛ/ and /æ/ are closer to one another in F1 x F2 space than /e/ and /a/ (the Spanish analogues), higher proficiency may reduce phonetic distance between the two vowels (resulting in a lower Pillai score). This would imply that higher proficiency speakers would have production categories more like the merged /ɛ-/æ/ category of their Dutch interlocutors (which also has a low Pillai score) than lower proficiency speakers. From a purely quantitative standpoint, then, higher proficiency speakers would have less phonetic distance to travel (here measured by change in Pillai score) to align to Dutch production patterns than speakers of lower proficiency groups. Since lower proficiency speakers would have more distance in F1 x F2 space to traverse to reach Dutch-like production benchmarks, we would reasonably expect the change in Pillai score across speech styles to be greater for lower proficiency groups. This would be evidenced by an interaction between proficiency and style, which we do not find. The lack of interaction between proficiency and the observed style shift for the /ɛ-/æ/ contrast does not provide support for an account based solely on a correlation between proficiency and more monolingual-like target production.

While alignment occurred to both vowel contrasts, it is crucial to note that the factors predicting alignment were distinct for each contrast. For /ɛ-/æ/, which involves alignment via merger, speech style and proficiency play a role in alignment, and the effect of proficiency is approximately quadratic. For /i-/ɪ/, which requires a phonological category split for alignment to occur, an interaction of Speech Style and Normalized Time is also present in addition to the main effects of Speech Style and Proficiency, which differentiates the latter contrast from the former. The findings suggest that alignment via merger is constituted by a rapid, general shift, which may be easier to produce than alignment via separation. The /ɛ-/æ/ contrast is more familiar to the Spanish speakers than to the Dutch confederates, and merging a distinction one can make is certainly simpler than creating a new one (see e.g., Flege et al., 2003; Escudero and Chládková, 2010). Consider also the finding in studies on language change that mergers are more common than splits historically (e.g., Campbell, 2013:19). Similar trends may apply to online processing of phonetic input, which would result in merger of /ɛ-/æ/ before separation of /i-/ɪ/. Our data suggest this is the case.

In general, we find that alignment to the interlocutor increases as cognitive processing load decreases. More proficient speakers can more easily inhibit activation of L1 phonological categories, allowing them to align to their interlocutor via merger in either speech style. When cognitive processing load is reduced in informal speech, the degree of merger also increases. Proficient speakers cannot, however, inhibit activation of the /i/ category in their L1, which occupies the F1 x F2 space of both /i/ and /ɪ/ in English, when the discourse context is more cognitively demanding (i.e., in formal speech). When the speech context is conducive to reducing monitoring and freeing cognitive resources for processing (i.e., in informal speech), there is a tendency to align by splitting the /i/ category into /i/ and /ɪ/ incrementally over time, but this is not modulated by proficiency. Importantly, it appears that the effects of speech style and proficiency are additive, but that speech style may be more critical for phonetic alignment (especially via split) than proficiency in the language of discourse in the current data.

One possible factor that has gone undiscussed so far is social motivation. In contrast to most accounts born from cognitive psychology (e.g., the Interactive Alignment Model; Pickering and Garrod, 2004), which posit that alignment is a mechanistic phenomenon primarily driven by lexical, phonological, and syntactic access during language processing, Communication Accommodation Theory (CAT; e.g., Giles et al., 1991), for example, would conclude that social

motivation is a chief mechanism underlying the observed trends. In a socially motivated framework such as CAT, alignment is postulated to occur to reduce social distance between a speaker and his/her interlocutor, which leads to a more positive attitude toward that speaker in the listener's mind and an improved social relationship between the interlocutors.

With the current dataset, we are not able to explicitly test an account based on social motivation, as the attitudes and personalities of the Spanish learners were not collected in the compilation of the corpus. However, it would stand to reason that the style shifting observed in the dataset is linked to social motivation and social distance. Under a sociolinguistic monitoring account of processing, the observed effect of style shifting would imply lower sociolinguistic monitoring in the informal situation. Decreased monitoring would concomitantly imply decreased cognitive processing load. These may work together to facilitate processing and lead to increased sensitivity to social cues (e.g., Labov et al., 2011), which might lead to a greater tendency to align to the interlocutor. However, although social motivations are undoubtedly important factors influencing one's production patterns toward an interlocutor, the current data do not permit us to disentangle the influences of cognitive load relative to social motivation; this would nonetheless be a fruitful area for future research.

The results of the current study have important implications for research on phonetic alignment and bilingualism. First, the findings suggest that while alignment seems to be a robust phenomenon in conversational interaction, its realization is influenced by the context in which speakers converse. Alignment is most likely to occur in informal discourse, where less cognitive resources are expended monitoring one's speech and where social pressures are reduced. Future research examining alignment should aim to create naturalistic, conversational environments in which to investigate phonetic alignment to maximize the likelihood that (a) alignment will occur and (b) that the findings are ecologically valid.

Second, our data also strongly suggest that the influence of time on alignment may vary with the phonological category in question. The underlying cross-linguistic interaction may produce distinct results: alignment may proceed either as a rapid shift or as the sum of incremental changes over time, depending on whether alignment requires the creation or the suspension of a phonological boundary relative to the L1 phonology. A temporally coarse-grained approach eschews this possibility. We therefore recommend that future research on phonetic alignment investigate how phonetic or phonological variables dynamically update over time rather than comparing pre- and post-exposure conditions.

Third, we found that alignment was modulated by phonological constraints, in this case represented as the influence of L1 categories in the participants' English production. When alignment to one's input requires the extemporaneous creation of a new phonological boundary, alignment appears to occur slowly over the course of the conversation (as exposure to tokens on either side of the category boundary increases). Multiple variables should therefore be used to quantify alignment, and we recommend that their selection be phonologically motivated.

The finding that the time course of alignment varies based on the L1 phonology raises the question of the importance of typological differences in the native languages of the interlocutors using a *lingua franca* in facilitating or inhibiting alignment. To our knowledge, Costa et al. (2008) were the first to suggest that a key factor modulating alignment among two non-native speakers would be the typological distance between their native languages. Costa et al. hypothesize that speakers with typologically distinct L1s will show reduced alignment in L2 discourse with one another (a

possible exception being speech rate; 2008, pp. 549-551). Dutch has a tense/lax distinction, front rounded vowels, and vowel reduction in unstressed syllables; none of these are commonplace in Spanish, which has only five vowels situated around the periphery of the vowel space. Based on these considerable typological differences at the level of phonology, Costa et al. would hypothesize that alignment in the current study should be minimal.

Despite the typological distance between Spanish and Dutch, we found alignment for both vowel contrasts. This is in agreement with recent work by Trofimovich and Kennedy (2014), who found evidence for discourse alignment in a study with speakers of English as a *lingua franca* from a variety of native-language backgrounds. This provides support for the claim that phonetic alignment will occur among interlocutors, independent of their language background. The findings from the current study combined with Costa et al.'s hypothesis would generate the prediction that alignment in phonological category production should be even greater among speakers of typologically similar languages.

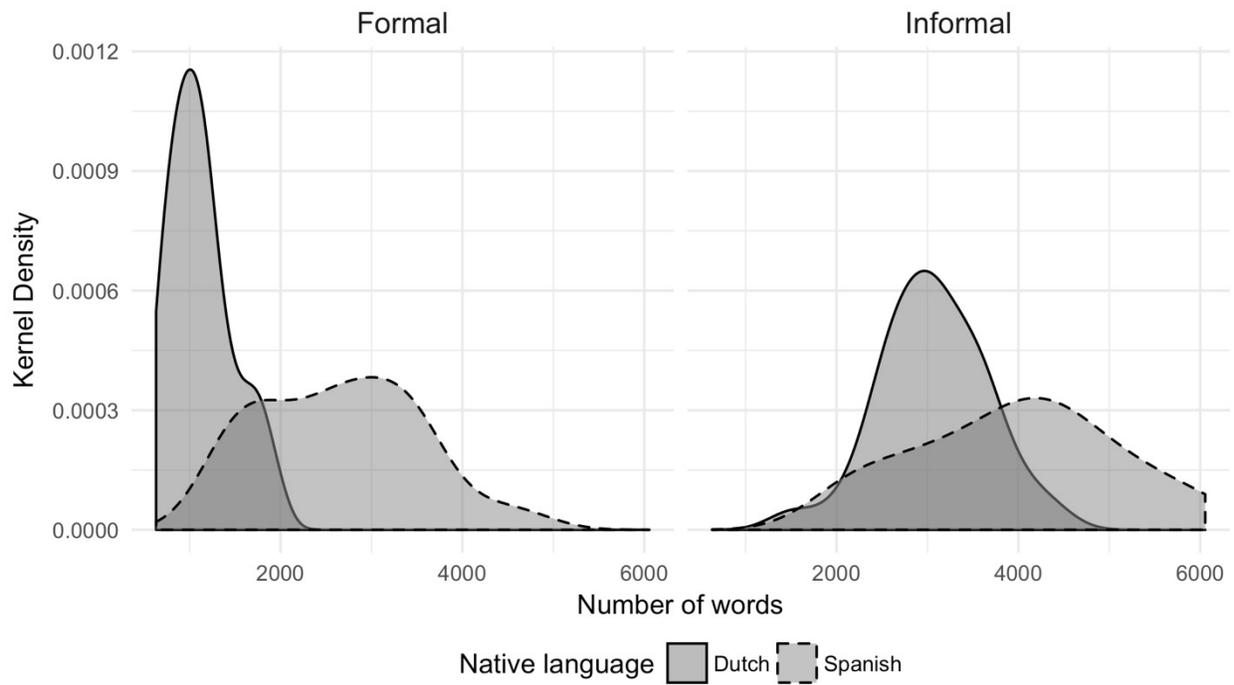
To develop a clearer portrait of how phonological category boundaries dynamically update in discourse in monolingual and bilingual populations alike, future research would benefit from an approach that links laboratory tasks on individual differences in the ability to translate perception to production (e.g., Kittredge and Dell, 2016) with naturalistic speech. Results from the current study suggest that an individual's availability of cognitive resources in language processing influences his or her degree of alignment in naturalistic discourse, and that cross-linguistic interference must be attenuated for alignment to occur. The cognitive processing strategies individuals utilize could be operationalized using measures of cognitive control (e.g., working memory, inhibition), which has previously been used to index phonological category competition within and across a bilingual's languages (e.g., Lev-Ari and Peperkamp, 2013, 2014; Moreno et al., 2008). One's strategic engagement of cognitive resources may also influence phonological alignment in a *lingua franca* (e.g., prepotent response inhibition, resistance to proactive interference, proactive and reactive control strategies; cf. Friedman and Miyake, 2014; Braver et al., 2007; Braver, 2012; Berry, 2016). These measures may be critical for understanding who is most likely to align, particularly in situations where cognitive resources are reduced due to monitoring (e.g., formal speech) or limited language access (e.g., proficiency).

Conclusion

The current study utilized unique aspects of bilingual phonology (namely, the transfer of L1 categories to the L2) to shed new light on phonetic alignment and the plasticity of bilingual phonological categories in discourse in a *lingua franca*. First, we investigated phonetic alignment between speakers in real time rather than post hoc, allowing us to document both a gradual (/i/-/ɪ/) and a rapid (/ɛ/-/æ/) process of alignment. Second, we analyzed variability in the production of carefully chosen vowel contrasts based on cross-language interference, which addresses a lacuna in current research. Previous studies have found discourse alignment to fine phonetic features within a single phonological category in monolingual populations, but we provide evidence that the boundaries across categories are themselves malleable in discourse, being shaped by input from one's conversational partner, the conversational context, and proficiency in the language of discourse.

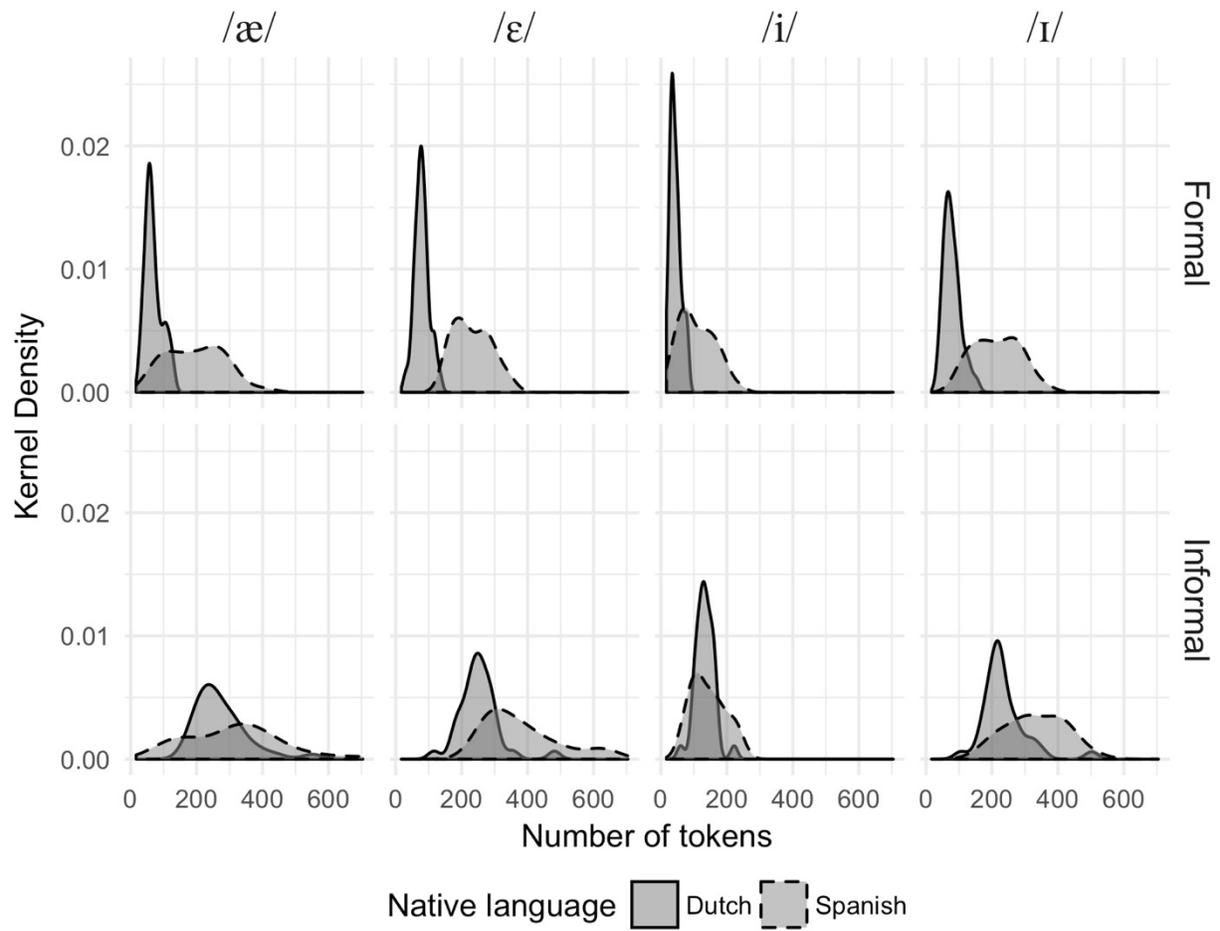
These findings together suggest that in the absence of strong biasing factors such as sociopolitical sentiment or language ideology, alignment is the norm rather than the exception. An individual's language history is projected in his or her speech, which is in turn perceived and refracted in the

output of his or her interlocutor through alignment. The consequence of this process is the propagation of variation within and across individuals, creating a perpetually changing phonological landscape whose topography remains to be adequately understood.

Appendix 1**Figure 5.** *Distribution of words spoken by native language and speech style across all conversations in the Nijmegen Corpus of Spanish English (NCSE; Kouwenhoven et al., 2015).*

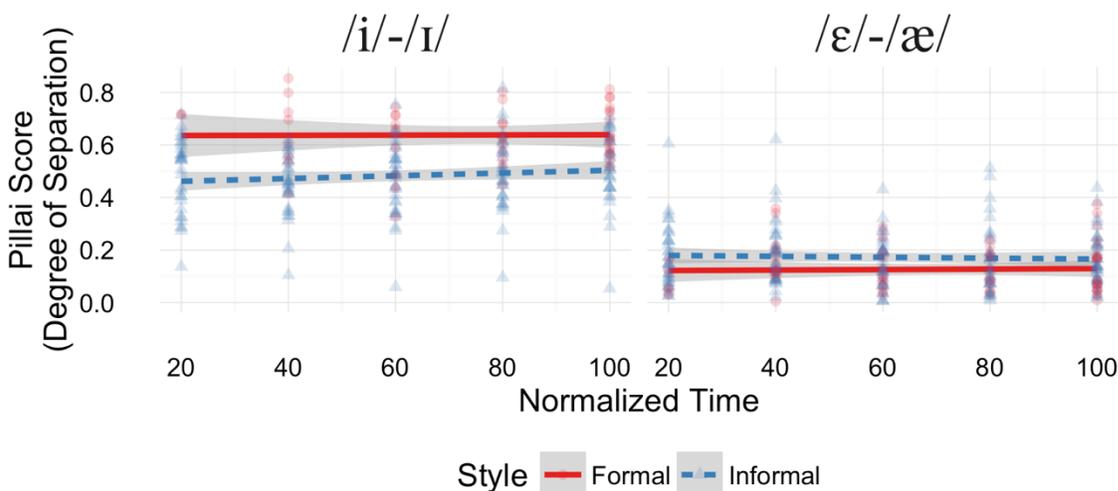
Appendix 2

Figure 6. *Distribution of extracted vowels from target contrasts by native language and speech style across all conversations.*



Appendix 3

Figure 7. Plot of Dutch Pillai scores by normalized time, speech style, and contrast.

Table 3. Mixed model predicting Dutch Pillai score for /i/-/ɪ/.^a

Predictor	β -Estimate	Std. Error	<i>t</i> -value
(Intercept)	0.59	0.02	28.95
Informal Style	-0.16	0.07	-2.24

^aRandom intercept by confederate sex, Random slope by speech style

Table 4. Mixed model predicting Dutch Pillai score for /ε/-/æ/.^a

Predictor	β -Estimate	Std. Error	<i>t</i> -value
(Intercept)	0.147	0.009	16.04

^aRandom intercept by confederate sex, Random slope by speech style

Notes

¹ Here we define a bilingual as anyone capable of sustaining conversation in multiple languages.

² There is evidence that Spanish native speakers assimilate Dutch /i/ and /ɪ/ to their native /i/ (Escudero and Williams, 2011; see also the work by Karin Wanrooij). Thus, even if Dutch speakers in the current study have distinct production targets relative to those of native English speakers, this is unlikely to facilitate perception of the contrast for the Spanish participants.

³ It is important to note that we are interested in the effect of speech register on alignment, not the effect of speaker sex. Speaker sex was one of multiple variables manipulated to influence register in the design of the NCSE corpus and thus cannot be analyzed independently of it. Any influence of speaker sex parity on alignment is necessarily also a component of the manipulation of speech style, as are attire and the belief that one is being video recorded. We encourage future studies to explore the relative influence of each of these factors independently, but this lies outside of the purview of the current study.

⁴ Note that the term ‘alignment’ in the acronym FAVE refers to the process of segmenting an audio file into words and phones based on a transcription given to the program (i.e., ‘aligning’ the transcribed words and their corresponding phones to the audio signal). This is distinct from the behavioral phenomenon described in this paper, in which a speaker sounds more like his/her interlocutor during or after conversational interaction.

⁵ The pronunciation models have also been found to be reliable for Standard Southern British English (MacKenzie and Turton, 2013), which is the variety most familiar to the participants and confederates in this study.

⁶ We adopted a normalized time approach in the current study to facilitate comparison of files with one another and more clearly present the results. To assuage concern regarding potential imbalances in tokens across time bins, a linear regression was performed on token counts by vowel, speech style, and native language. We found no main effect of time bin, no two-way interaction between native language and time bin, and no three-way interaction among native language, time bin, and speech style.

⁷ A density plot of the token distributions by time bin, vowel contrast, speech style, and native language is available in the online supplemental materials (TimeBin_counts.tiff), accessible via the Open Science Framework: <https://osf.io/ut86w/>.

⁸ Data files and R code are available in the online supplemental materials.

⁹ All eigenvalues must be positive, since the Pillai score is based on products of positive definite matrices. For this case, where $p = 1$, it must hence be true that $0 < \lambda/(1 + \lambda) < 1$.

¹⁰ The fully specified model for the confederates was like all other Pillai models in terms of fixed effects structure, with exception that Proficiency was removed as a fixed effect. The random effects structure included solely a random intercept by confederate sex, which in this case is equivalent to a by-subject intercept. The model reduction procedure was otherwise identical to the procedure outlined in Section 1.6.

¹¹ Vowel data by native language and target contrast (span_i.csv, span_e.csv, dutch_i.csv, dutch_e.csv) have been included in the online supplemental materials, as has the script used to calculate Pillai scores and analyze the data (Pillai_analysis.R).

References

- Adank, P, Smits, R and van Hout, R. (2004a). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116, 3099-3107.
- Adank, P, van Hout, R and Smits, R. (2004b). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America*, 116, 1729-1738.
- Adank, P, van Hout, R and van de Velde, H. (2007). An acoustic description of the vowels of northern and southern standard Dutch II: regional varieties. *Journal of the Acoustical Society of America*, 121, 1130–1141.
- Archila-Suerte, P, Zevin, J, Bunta, F and Hernández, A.E. (2012). Age of acquisition and proficiency independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, 15: 190-201.
- Baayen, RH, Davidson, D and Bates, D. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59: 390-412.
- Baayen, RH, Kliegl, R and Vasishth, S. (2015). RePsychLing: Data sets from Psychology and Linguistics experiments. R package version 0.0.4.
- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society* 39: 437–456.
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40: 177–189.
- Babel, M and Bulatov, D. (2011). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech* 55: 231–248.
- Babel, M, McAuliffe, M and Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology* 4: 321–323.
- Barr, DJ, Levy, R, Scheepers, C and Tilly, HJ. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68: 255-278.
- Barrios, S, Jiang, N and Idsardi, WJ. (2016). Similarity in L2 phonology: Evidence from L1 Spanish late learners' perception and lexical representation of English vowel contrasts. *Journal of the Acoustical Society of America*, 129: EL(1).
- Baese-Berk, MM, Bradlow, AR and Wright, BA. (2013). Accent-independent adaptation to foreign-accented speech. *JASA Express Letters, Journal of the Acoustical Society of America*, 133, EL 174.
- Bates, D, Maechler, M, Bolker, B and Walker, S. (2015a). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67: 1-48.
- Bates, D, Kliegl, R, Vasishth, S. and Baayen, RH. (2015b). Parsimonious mixed models. Arxiv.org: <http://arxiv.org/pdf/1506.04967v1.pdf>
- Berry, GM. (2016). Processing linguistic variation through dual mechanisms of cognitive control. *Linguistics Vanguard*, 2(s1): 85-96.
- Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8: 9-37.

- Biber, D, Davies, M, Jones, JK and Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora*, 1(1): 1-37.
- Bohn, OS and Flege, JE. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11: 303-328.
- Bradlow, AR. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97: 1916-1924.
- Bradlow, AR and Alexander, JA. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121: 2339-2349.
- Bradlow, AR and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106: 707-729.
- Branigan, HP, Pickering, MJ and Cleland, AA. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75: B13-25.
- Braver, TS. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Science*, 16(2) : 106-113.
- Braver, TS and Barch, DM. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience and Biobehavioral Reviews* 26: 809-817.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
- Burgos, P, Cucchiaroni, C, van Hout, R and Strik, H. (2013). Pronunciation errors by Spanish learners of Dutch: A data-driven study for ASR-based pronunciation training. *Proceedings of Interspeech 2013*: 2385-2389.
- Campbell, L. (2013). *Historical linguistics: An introduction*. 3rd edition. Edinburgh: Edinburgh University Press.
- Casillas, JV. (2015). Production and perception of the /i/-/ɪ/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica*, 72: 182-205.
- Casillas, JV and Simonet, M. Production and perception of the English /æ/-/ɑ/ contrast in switched-dominance speakers. *Second Language Research*, 32(2): 171-195.
- Chang, CB. (2012). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41: 520-533.
- Chang, CB and Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *Journal of the Acoustical Society of America*, 132: 2700-2710.
- Chartrand, TL, and Bargh, JA. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893-910.
- Clopper, CG. (2014). Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology*, 5(1): 69-90.
- Costa, A, Pickering, MJ and Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, 23: 528-556.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

- Davis, MH, Johnsrude, IS, Hervais-Adelman, A, Taylor, K and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded segments. *Journal of Experimental Psychology: General*, 134: 222-241.
- de Looze, C, Scherer, S, Vaughan, B and Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58: 11-34.
- Delvaux, V and Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64: 145-173.
- Elston-Güttler, KE, Paulmann, S and Kotz, SA. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, 17(10): 1593-1610.
- Ernestus, M, Hanique, I and Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics*, 48: 60-75.
- Escudero, P and Chládková, K. (2010). *Journal of the Acoustical Society of America*, 128(5): EL254-EL260.
- Escudero, P and Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *Journal of the Acoustical Society of America*, 129(1): EL1-EL7.
- Escudero, P, Simon, E and Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, 40, 280-288.
- Flege, JE. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology*, 43A: 701-731.
- Flege, JE, Bohn, OS and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25: 437-440.
- Flege, JE, Schirru, C and MacKay, IRA. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40: 467-491.
- Friedman, NP and Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1) : 101-135.
- Ganong, WF. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6: 110-125.
- Garcia Lecumberri, ML, Cooke, M and Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, 52: 864-886.
- Giacomino, L. (2012). Comparative analysis of vowel space of L1 Spanish speakers and General American English. *Linguistic Portfolios*, 1, Article 9.
http://repository.stcloudstate.edu/stcloud_ling/vol1/iss1/9
- Giles, H, Coupland, N and Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In Ekman, P., and Scherer, K.R.

- (eds.), *Contexts of Accommodation*, 1–68. New York: Cambridge University Press.
- Gussenhoven, C. (1992). Dutch (Illustrations of the IPA). *Journal of the International Phonetic Association*, 22: 45-47.
- Hay, J, Warren, P and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34: 458–484.
- Hay, J and Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48: 865-892.
- Hillenbrand, J, Getty, LA, Clark, MJ and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5): 3099-3111.
- Hwang, J, Brennan, SE and Huffman, MK. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, 81: 72–90.
- Jacobs, CL, Yiu, LK, Watson, DG and Dell, GS. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language*, 84: 37-48.
- Jesse, A and McQueen, JM. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin and Review*, 18: 943-950.
- Johnson, K. (2005). Speaker Normalization in speech perception. In D. B. Pisoni and R. Remez (eds.) *The Handbook of Speech Perception*. Oxford: Blackwell: 363-389.
- Kim, M, Horton, WS and Bradlow, AR. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology*, 2: 125-156.
- Kittredge, AK and Dell, GS. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, 89: 8-22.
- Kondaurova, MV, and Francis, AL. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense-lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, 124(6): 3959-3971.
- Kouwenhoven, H, Ernestus, M and van Mulken, M. (2015). Register variation by Spanish users of English: The Nijmegen Corpus of Spanish English. *Corpus Linguistics and Linguistic Theory*. Epub ahead of print 10 Apr 2015. doi:10.1515/cllt-2013-0054.
- Kraljic, T and Samuel, AG. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51: 141-178.
- Kraljic, T and Samuel, AG. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56: 1-15.
- Kraljic, T, Brennan, SE and Samuel, AG. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107: 54-81.
- Labov, W. (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Malden, MA: Blackwell.

- Labov, W, Ash, S, Ravindranath, M, Weldon, T, Baranowski, M and Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4): 431-463.
- Lev-Ari, S and Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, 41: 320-331.
- Lev-Ari, S and Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, 47: 36-46.
- Levelt, WJM. (1983). Monitoring and self-repair in speech. *Cognition*, 14: 41-104.
- Levey, S and Cruz, D. (2004). The discrimination of English vowels by bilingual Spanish/English and monolingual English speakers. *Contemporary Issues in Communication Science and Disorders*, 31: 162-172.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the HandH theory. In W. J. Hardcastle and Marchal, A. (eds.). *Speech Production and Speech Modeling*. Kluwer: Dordrecht: 403-439.
- Mackenzie, L and Turton, D. (2013). Crossing the pond: Extending automatic alignment techniques to British English dialect data. *New Ways of Analyzing Variation (N WAV)* 42 (October).
- Magnuson, JS and Nusbaum, HC. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33: 391-409.
- Martínez-Celdrán, E, Fernández-Planas, AM and Carrera-Sabaté, J. (2003). Castilian Spanish (Illustrations of the IPA). *Journal of the International Phonetic Association*, 33: 255-259.
- Mattys, SL and Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65: 145-160.
- Mattys, SL, Barden, K and Samuel, AG. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin and Review*, 21: 748-754.
- Mattys, SL and Palmer, SD. (2015). Divided attention disrupts perceptual encoding during speech recognition. *Journal of the Acoustical Society of America*, 137: 1464.
- Maye, J, Aslin, RN and Tanenhaus, MK. (2008). The Weckud Wetch of the West: Lexical adaptation to a novel accent. *Cognitive Science*, 32: 543-562.
- Mitterer, H and Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109: 168-173.
- Moreno, EM., Rodríguez-Fornells, A and Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *Journal of Neurolinguistics*, 21: 477-508.
- Namy, LL, Nygaard, LC, and Sautereig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21: 422-432.
- Nielsen, K. (2011). Specificity and abstractness in VOT imitation. *Journal of Phonetics*, 39: 132-142.

- Nycz, J and Hall-Lew, L. (2013). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, 20. 060008. doi:10.1121/1.4894063.
- Olson, CL. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83: 579-586.
- Pardo, JS. (2001). Imitation and coordination in spoken communication. PhD Thesis. Yale University.
- Pardo, JS. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119: 2382.
- Pickering, MJ and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27: 169-225.
- Pillai, KCS. (1954). On some distribution problems in multivariate analysis. Special report to the Ford Foundation. *Institute of Statistics Mimeograph Series*, Number 88.
- Pillai, KCS. (1955). Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics*, 26: 117-121.
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 3.3.0. URL <https://www.R-project.org/>
- Rosenfelder, I, Fruehwald, J, Evanini, K, Seyfarth, S, Gorman, K, Prichard, H and Yuan, J. (2014). FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2 10.5281/zenodo.22281
- Sancier, ML and Fowler, CA. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25: 421-436.
- Sharifzadeh, HR, McLoughlin, IV and Russell, ML. (2012). A comprehensive vowel space for whispered speech. *Journal of Voice*, 26(2): e49-e56.
- Shockely, K, Sabadini, L and Fowler, CA. (2004). Imitation in shadowing words. *Perception and Psychophysics*, 66: 422-429.
- Sidasas, SK. (2011). Hearing what you expect to hear: The interaction of social and cognitive mechanisms underlying vocal accommodation. PhD Thesis. Emory University.
- Simon, E and D'Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast. *Language Sciences*, 34: 269-283.
- Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, 43: 26-37.
- Thompson, GL and Brown, AV. (2012). Interlanguage variation: The influence of monitoring and contextualization on L2 phonological production. *Vigo International Journal of Applied Linguistics*, 9: 107-132.
- Trofimovich, P and Kennedy, S. (2014). Interactive alignment between bilingual interlocutors: Evidence from two information-exchange tasks. *Bilingualism: Language and Cognition*, 17: 822-836.
- Trudgill, P. (1986). *Dialects in Contact (Language in Society)*. Oxford: Blackwell.

- van Engen, KJ and Bradlow, AR. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, 121: 519-526.
- Vroomen, J, van Linden, S, de Gelder, B and Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45: 572-577.
- Wang, H and van Hueven, VJ. (2006). Acoustic analysis of English vowels produced by Chinese, Dutch, and American speakers. *Linguistics in the Netherlands*, 23: 237-248.
- Weber, A, Di Betta, AM and McQueen, JM. (2014). Treack or trit: Adaptation to genuinely and arbitrary foreign accents by monolingual and bilingual learners. *Journal of Phonetics*, 46: 34-51.
- Weinreich, U. (1968). *Languages in Contact: Findings and Problems*. The Hague: Mouton.
- White, EJ, Titone, D, Genesee, F and Steinhauer, K. (2017) Phonological processing in late second language learners: The effects of proficiency and task. *Bilingualism: Language and Cognition*. 20(1): 162-183. doi: 10.1017/S1366728915000620.
- Wilson, I and Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, 57: 361-373.
- Witteman, MJ, Bardhan, NP, Weber, A and McQueen, JM. (2015). Automaticity and stability of adaptation to a foreign-accented talker. *Language and Speech*, 58(2): 168-189.