Vocal alignment to native and non-native speakers of English

Eva M. Lewandowski^{a)} and Lynne C. Nygaard

Department of Psychology, Emory University, 36 Eagle Row, Atlanta, Georgia 30322, USA

(Received 30 May 2017; revised 1 May 2018; accepted 2 May 2018; published online 8 August 2018)

Research on vocal alignment, the tendency for language users to match another individual's speech productions, suggests that multiple factors contribute to this behavior. Social and motivational goals, aspects of cognitive architecture, and linguistic flexibility may all affect the extent to which vocal alignment occurs, suggesting complex underlying mechanisms. The present study capitalized on the social and linguistic characteristics of Spanish-accented English to examine the relationship among these contributors to vocal alignment. American English-speaking adults participated in a shadowing task. Degree of vocal alignment was assessed by both acoustic measures and independent raters' judgments. Participants aligned to both native English and Spanish-accented productions, despite differences in attitudes to and intelligibility of the different accents. Individual differences in shadowers' vowel dispersion were also related to extent of vocal alignment, with greater dispersion associated with greater alignment. Acoustic measures were related to perceptual assessments of alignment and differed by accent type, suggesting that patterns of alignment may differ across accents. Overall, the current study demonstrates vocal alignment between talkers of differing language backgrounds and highlights the importance of acoustic and linguistic components of alignment behavior. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5038567>

[TB]

Pages: 620–633

I. INTRODUCTION: VOCAL ALIGNMENT TO NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH

Speech is a highly variable signal influenced by social, linguistic, individual, and situational factors (Bradlow and Pisoni, 1999; Goldinger *et al.*, 1991; Nygaard and Pisoni, 1998; Peterson and Barney, 1952). Language users are perceptually sensitive to variations in speech, and will *vocally align* their own productions to more closely match other talkers' speech patterns (e.g., Bilous and Krauss, 1988; Pardo, 2006). Vocal alignment is defined here as convergence between two language users along any linguistic dimension (e.g., Pickering and Garrod, 2004; Pardo, 2006). Although many studies have examined the social and linguistic factors that influence individuals' tendency to engage in vocal alignment (e.g., Gregory and Webster, 1996; Pardo *et al.*, 2010), the underlying causes remain poorly understood. The present study examined vocal alignment to non-native speakers to explore the social, perceptual, and cognitive mechanisms that underlie vocal alignment.

A. Theoretical accounts of vocal alignment

One class of theoretical accounts assumes that social factors primarily drive alignment behavior. Specifically, communication accommodation theory (Giles *et al.*, 1991; Giles and Ogay, 2007) posits that alignment's primary function is to modulate social distance (Bourhis and Giles, 1977). Supporting this hypothesis, research has shown that interlocutors' attitudes and status influence degree of vocal alignment (Abrego-Collier *et al.*, 2011; Bilous and Krauss, 1988; Namy *et al.*, 2002). Positive attitudes towards a dialect

group (Babel, 2010) or voice (Babel *et al.*, 2014) as well as attractiveness (Babel, 2012) promote vocal alignment. Individuals with high status tend to receive more alignment than individuals with lower status (Bourhis and Giles, 1977; Gregory and Webster, 1996; Giles, 1973; Pardo, 2006). Gender may also interact with other social-motivational factors to condition alignment behavior (Namy *et al.*, 2002; but see Pardo *et al.*, 2017). Taken together, these studies suggest that positive attitudes and social preferences increase alignment behavior.

Social and motivational factors may not necessarily be the only or primary mechanisms driving alignment behavior. In laboratory shadowing tasks, participants reliably align to voices on a variety of measures even without interpersonal contact (Babel *et al.*, 2014; Dias and Rosenblum, 2016; Fowler *et al.*, 2003; Goldinger, 1998; Kim, 2012; Mitterer and Müsseler, 2013; Namy *et al.*, 2002; Pardo *et al.*, 2013; Shockley *et al.*, 2004). The presence of vocal alignment without an explicit conversational setting highlights the contribution of stimulus-driven properties, both linguistic and non-linguistic, to alignment behavior, perhaps reflecting the operation of low-level perceptual-motor systems.

A number of proposals posit that alignment relies on a fundamental (i.e., low-level) relationship between speech perception and speech production. These proposals suggest that vocal alignment may increase mutual intelligibility among talkers by aligning representations to the detailed pronunciation differences across individuals and groups (Fowler and Galantucci, 2005; Liberman and Mattingly, 1985; Pickering and Garrod, 2004; Garrod and Pickering, 2013; Shockley *et al.*, 2004). By such accounts, perceptual exposure to variations in the realization of linguistic form can alter or prime changes in the production of speech

^{a)}Electronic mail: elewand@emory.edu

(Sancier and Fowler, 1997; Munro *et al.*, 1999) because perception and production are fundamentally linked. For example, Pickering and Garrod (2004, 2013) propose that priming at multiple linguistic levels enables language users to align representations and generate predictions about the form of upcoming utterances, thus increasing mutual intelligibility. The perception-production link has also been conceptualized as part of a general action understanding system (Fowler *et al.*, 2003; Prinz, 1997) in which productive mechanisms are recruited for understanding phonetic form. Supporting this proposal, Adank *et al.* (2010) found that imitating a novel Dutch accent resulted in greater gains in intelligibility than other kinds of perceptual training. In general, these types of accounts suggest that input to the speech perception system may necessarily affect speech production because perception and action are tightly related.

Although both social and perceptual-motor factors appear to influence alignment behavior, the specific mechanisms that support the operation and interaction of these factors are unclear. Proponents of the social-motivational approach are often agnostic as to whether alignment processes are intentional and motivated or implicit and automatic (Dragojevic *et al.*, 2015). Likewise, accounts focusing on perception-production correspondences do not fully explain how social factors may influence low-level alignment processes (Babel, 2012; Pardo, 2006; Pickering and Garrod, 2013). In order to understand how social-motivational and perceptual-motor mechanisms determine degree of vocal alignment, the current study examined alignment to non-native speakers. The linguistic and social properties associated with non-native speakers provide an informative test case for assessing how social attitudes and perceptual properties of speech shape the mechanisms underlying vocal alignment.

B. Vocal alignment to non-native speech

As a social index, non-native accents convey a host of information to a native listener, including the talker's status as a non-native speaker and other social factors accompanying language use (Atagi and Bent, 2015, 2017; Giles and Ogay, 2007; Gluszek and Dovidio, 2010). Young children will choose to befriend individuals with their own accent over non-native or foreign-language speaking individuals, and children's same-dialect preference is more powerful than their same-race preference (Kinzler *et al.*, 2009). Non-native speech is also often associated with more negative attitudes than standard speech (McKirnan and Hamayan, 1984; Ryan, 1983; Park-Taylor *et al.*, 2008). Native English speakers rate Spanish-English bilinguals as having lower status and competence and as less likely to make friends than native speakers (Nelson *et al.*, 2016).

As a perceptual signal, non-native speech differs from and is often less intelligible than native speech (Bradlow and Bent, 2008; Sidaras *et al.*, 2009; Sumner, 2011). Differences in lexical stress, intonation, and phonetic category structure (Munro and Derwing, 1995) contribute to differences in intelligibility between non-native and native speech. The decreased intelligibility of non-native speakers can in turn

influence attitudes toward the speakers (Bresnahan *et al.*, 2002; Dragojevic and Giles, 2016). In sum, previous research indicates that non-native speech is associated with more negative social attitudes and is less intelligible than native speech.

The social and perceptual characteristics of non-native speech have important implications for our understanding of the causal underpinnings of vocal alignment. Although there is evidence that both social and perceptual-motor factors influence vocal alignment, non-native speech permits an examination of the relative weighting of these factors, leading to different predictions with respect to vocal alignment. Although research has shown that vocal alignment occurs across dialect groups (Delvaux and Soquet, 2007; Walker and Campbell-Kibler, 2015) and can be modulated by positive or negative social attitudes (e.g., Babel, 2010), few studies have explored alignment toward non-native speakers as a function of both the speakers' intelligibility and attitudes toward non-native speech (Kim, 2011, 2012). Kim *et al.* (2011) examined vocal alignment of dyads matched or mismatched on language background. Overall, the authors found that vocal alignment in the matched dyads was greater than in mixed language dyads. The authors proposed that partners of similar language backgrounds engage in greater alignment because non-native pronunciations may be too far outside the native speaker's articulatory and representational structure (Kim *et al.*, 2011). However, the lack of reliable vocal alignment to non-native speakers may stem from a host of other factors, including intelligibility of the non-native talkers, participant familiarity with non-native speech, and attitudes toward non-native speakers.

C. The current study

The primary goal of this experiment was to investigate the relative contributions of perceptual and social characteristics to vocal alignment by examining native English speakers' alignment to Spanish-accented speakers in a laboratory shadowing task (e.g., Goldinger, 1998). In addition to the differences in attitudes toward and intelligibility of Spanish-accented and native speech, there are well-characterized differences in vowel inventory between Spanish and English (e.g., Bradlow, 1995) that are reflected in Spanish-accented English (e.g., Sidaras *et al.*, 2009). These differences permit a clear examination of the role of linguistic distance in vocal alignment (Kim *et al.*, 2011). Because contrasting patterns of alignment would be expected for the particular attitudinal and acoustic-perceptual properties associated with Spanish-accented English, Spanish-accented speech affords the opportunity to disambiguate the relative contributions of social or perceptual attributes in guiding alignment behavior.

Theoretical accounts of vocal alignment lead to contradictory predictions about the extent to which native English speakers will vocally align to Spanish-accented speakers. Shadowers would be expected to align less to speakers that are judged less positively (e.g., lower status; Babel, 2012; Bourhis and Giles, 1977; Gregory and Webster, 1996) if social attitudes are weighted more heavily in vocal alignment processes. Therefore, individuals should vocally align

less to Spanish-accented speakers than to native speakers because the former is typically associated with more negative attitudes. Less alignment would also be expected if non-native speech productions lie too far outside of the native production repertoire (Kim *et al.*, 2011). However, if lower intelligibility and greater acoustic differences between the target and the shadower are weighted more heavily in alignment, shadowers should align *more* to the less intelligible, more distinctive Spanish-accented speakers. This could facilitate mutual intelligibility through representational parity or priming of linguistic form (e.g., Adank *et al.*, 2010; Pickering and Garrod, 2004) or result from heightened “salience” of unfamiliar Spanish-accented speech stemming from violated expectations based on previous experience (Babel *et al.*, 2014; Goldinger, 1998; Jaeger and Weatherholtz, 2016).

A second goal was to examine the acoustic dimensions along which speakers align to native and non-native speakers and the degree of agreement between acoustic and perceptual measures. Assessments of vocal alignment included both listener judgments and acoustic measures. For a holistic measure of vocal alignment, an independent group of listeners rated similarity of the shadowers’ utterances to the model’s utterances in an AXB rating task (Goldinger, 1998; Miller *et al.*, 2013; Namy *et al.*, 2002; Pardo *et al.*, 2013). Although the AXB task is useful for determining global alignment, it does not identify specific dimensions along which alignment has occurred. As such, vocal alignment was also assessed by the acoustic measurements of fundamental frequency (f0), duration, and vowel spectra (e.g., Babel and Bulatov, 2012; Babel, 2010; Kim, 2011, 2012; Staum-Casasanto *et al.*, 2010) based on their use in past work.

A third goal of the study was to examine inter-individual differences in the propensity to vocally align. Although researchers have observed large differences between individuals in vocal alignment (Kim *et al.*, 2011; Pardo *et al.*, 2013; Pardo, 2013), accounting for these individual differences has been difficult (Pardo *et al.*, 2017; Yu *et al.*, 2013). For alignment to native speakers, greater phonetic distance between interlocutors may promote alignment (Babel, 2010, 2012). However, native English speakers may be less likely to align to non-native speakers because the non-native phonetic repertoire is outside the native range (Kim *et al.*, 2011). Although objective distance between the target and shadower may play a role, some researchers (Babel, 2010, 2012; Walker and Campbell-Kibler, 2015) also suggest that an individual’s baseline (e.g., typical) production repertoire may constrain the effect of phonetic distance on alignment. That is, an individual may be most likely to align when differences between the model talker and shadower are large, but only if the model talkers’ productions are within the shadower’s natural phonetic range. To assess this possibility, the present study examined whether intra-speaker variation in speech production relates to vocal alignment.

In summary, the current study addressed the following research questions: (1) Do native English-speaking shadowers differentially align to native Spanish speakers in comparison to native speakers of English on (a) perceptual and

(b) acoustic measures of vocal alignment? (2) Along what acoustic dimensions do speakers align and how do acoustic measures relate to perceptual measures of alignment? (3) Does variation or flexibility in shadower’s habitual speech productions relate to the extent of alignment and does this differ depending on characteristics of the model talker?

II. METHODS

A. Participants

Independent groups of participants recruited from the Emory University Psychology Department subject pool completed the shadowing task and ratings tasks described below. All participants were 18–25 years old, were native speakers of Standard American English (SAE; any exceptions are noted in the following), and reported no history of speech or hearing disorders. Participants received course credit as compensation.

1. Shadowers

Thirty-three participants (26 female) completed the shadowing task. Ten were excluded on the basis of language background (e.g., non-native English speaker; 2 shadowers) or computer technical error (8 shadowers). The remaining 23 shadowers (17 female) were native SAE speakers. Shadowers self-reported their familiarity with non-native accents, including Spanish-accented English. Five shadowers reported that they had a family member or close friend with a Spanish accent, but only two of those five reported exposure to the accent since starting college. Overall, self-reports indicated that the shadowers lacked extensive experience with Spanish-accented English.

2. AXB raters

One hundred and forty-four independent raters (117 female) completed the AXB judgment task. Eight were excluded due to computer technical error, and data were analyzed for the remaining 136 (110 female). Nine raters were bilingual [L2 = Chinese (2), Creole—unspecified type (1), German (1), Gujarati (1), Spanish (2), Telugu (1), Urdu (1)]. Bilingual raters were distributed across shadowers; no more than 1 (or 2) bilingual raters provided similarity judgments per shadower. The ratio of raters to shadowers was 5 (or 6) to 1. This number was chosen because recent work suggests that alignment ratings obtained by five raters per shadower are comparable in reliability to those obtained with ten raters per shadower (Pardo *et al.*, 2014; Pardo *et al.*, 2017).

B. Stimuli

Four talkers (2 female) selected from the Speech and Language Perception Lab’s databases served as models in the shadowing task: English male (EM) and English female (EF) as native targets (L1 = Standard American English), and Spanish male (SM) and Spanish female (SF) as non-native targets (L1 = Mexico City Spanish). SM arrived in the United States at age 22, learned English at age 20, and had lived in the United States for 7 years at time of recording. SF

arrived in the United States at 34 years old, began learning English at 2, and had been in the United States 3 years at the time of recording. The four model talkers were recorded reading 150 words and 100 sentences (see Sidasar *et al.*, 2009). From these materials, 72 low frequency ($\bar{x} = 17.66$; Brysbaert and New, 2009) CVC (consonant-vowel-consonant) words were chosen (see Appendix A). The word list included three diphthongal (/ou/, /ei/, /ai/) and eight monophthongal (/a/, /æ/, /i/, /ɪ/, /ɛ/, /u/, /ʌ/, /ɔ/) vowels. Vowel categories were represented unequally in the word set with the following number of observations per category: 11 /æ/, 10 /ɪ/, 10 /ʌ/, 9 /ei/, 8 /ou/, 6 /a/, 6 /i/, 4 /u/, 3 /ɛ/, 3 /ai/, 2 /ɔ/. Summary statistics of the model talkers' acoustic properties is provided in the supplementary material (Fig. S1 and Table S1).¹ Speech recordings were digitized at 22.050 kHz, amplitude normalized, and segmented into word-length files. These utterances were used as target stimuli in the shadowing task and as AXB comparison stimuli. The procedure for recording shadowers' utterances for use in the AXB task was identical to the one described for model talkers' utterances.

1. Intelligibility and accentedness measures

Thirty-two independent participants provided accentedness ($n = 16$; 11 female) and intelligibility measures ($n = 16$; 14 female) for the full set of stimuli from each model talker. Accentedness was measured on a 7-point Likert scale (1 = "not accented"; 7 = "very accented") and intelligibility was based on word transcription accuracy. Each word was presented once to each participant in a random order (e.g., not blocked by speaker or accent). The intelligibility and accentedness tasks confirm that our Spanish-accented stimuli were less intelligible [$t(15) = -16.16$, $p < 0.01$] and perceived as more accented [$t(15) = 16.86$, $p < 0.01$] than our native stimuli (see Table I). SF is more intelligible [$t(15) = 2.64$, $p = 0.01$] and rated as more accented [$t(15) = 3.48$, $p = 0.003$] than SM. The statistical relationships survive correction for multiple comparisons (corrected α -level = 0.0253; Šidák, 1967).

2. Attitudes ratings

An attitudes rating task was conducted to confirm attitude differences for the current speakers in our subject population. Sixteen participants (12 female) provided judgments on a 1 to 7 Likert-type scale for 21 attribute pairs (e.g.,

rich–poor; mean–nice) for each talker (Heaton and Nygaard, 2011). Listeners heard 18 words spoken by a single talker, then made their ratings. The attributes were reduced to two categories, "sociality" and "status," on the basis of a factor analysis (see Appendix B). The Spanish-accented talkers were perceived as lower status than [$t(15) = -3.14$, $p < 0.01$] but equally sociable [$t(15) = 0.67$, $p = 0.67$] to the native talkers (see Table I). Although the females appear to be rated higher as higher status than males [$t(15) = 2.26$, $p = 0.039$], the effects do not survive correction for multiple comparisons (Šidák-corrected $\alpha = 0.0256$). There are no significant differences among ratings of sociality for male versus female model talkers [$t(15) = 1.83$, $p = 0.087$]. The attitudes data confirm that listeners were not only able to form opinions about talkers based on word-length stimuli, but also that status judgments differed based on talker's accent.

C. Procedure

1. Shadowing task

The shadowing task consisted of three trial types presented in a fixed order: baseline, perceptual exposure, and shadowing (Goldinger, 1998). During baseline trials, participants saw each of the 72 stimulus words presented on a computer monitor and read words aloud as quickly and as clearly as possible. Participants read the word list twice in two randomized blocks to increase the probability of collecting usable word tokens. In the perceptual exposure and shadowing blocks, participants heard model talkers produce a unique subset of 18 words (e.g., "bead" spoken only by SM), forming model-word pairs. Model-word pairings were counterbalanced across shadowers with a Latin-square design.

Perceptual exposure occurred after the final baseline block and before the first shadowing block. As in the original shadowing paradigm of Goldinger (1998), the perceptual exposure block was included to familiarize the shadowers with all the model talkers' voices, as the same model talkers and words were presented during both exposure and shadowing trials. Although the exposure phase also included the utterances from the native talkers, exposure to the Spanish-accented speakers was particularly important, given the increased likelihood of naming errors (Bradlow and Bent, 2008; Sidasar *et al.*, 2009). During perceptual exposure, participants listened to 18 words produced by each model talker (72 total) and identified the word from an on-screen 3×3 table of choices. Alternatives were other items from the word list. Participants heard four repetitions of each word-model talker pair during the perceptual exposure block. All 72 words were presented before an item repeated. No feedback was given.

In the shadowing trials, participants said aloud words spoken by the model talkers. Task instructions specifically excluded words like "imitate" and "repeat" to avoid biasing participant responses. The full word list was presented twice, and order was randomized on each presentation. Trials in both the shadowing and perceptual exposure blocks were fully randomized (i.e., not blocked by speaker or accent).

The task was administered on a PC computer using E-prime software (Schneider *et al.*, 2002) in one testing

TABLE I. Intelligibility, accentedness, and attitudes ratings of the model talker stimuli. Intelligibility is measured as transcription accuracy. Mean accentedness scores based on a 1–7 Likert type rating, where higher numbers indicate greater degree of foreign accentedness. Sociality and Status composite Likert ratings of the 21 attributes collapsed into categories based on the results of factor analysis (Appendix B).

Model talker	Baseline intelligibility (% correct)	Accentedness rating (1–7 scale)	Sociality rating (1–7 scale)	Status rating (1–7 scale)
English Female	94	1.89	4.43	5.17
English Male	93	1.92	4.23	4.10
Spanish Female	66	5.63	4.69	4.08
Spanish Male	54	4.77	3.97	3.84

session. Shadower responses were recorded in Audacity using an Audio-Technica ATR20 cardioid low-impedance microphone. Participants were tested individually in sound-attenuated rooms and wore Beyerdynamic DT-100 headphones throughout the entire task, even when no auditory stimuli were presented.

2. AXB task

Because shadowers align on a variety of dimensions that may not be fully captured by individual acoustic characteristics (e.g., fundamental frequency), an AXB task was used to obtain a global measure of vocal alignment (Goldinger, 1998; Pardo *et al.*, 2013). An independent group of raters compared each model talker's production of an utterance (X) to the shadower's shadowed production (A) and to the shadower's baseline production (B). Stimulus order within a trial (AXB and BXA) occurred twice for each word and trials were presented in random order. If vocal alignment has occurred, raters should report a greater similarity between the shadowed production (A) and the model talker's production (X) than between the shadower's baseline utterance (B) and the model talker's production (X) at rates that are reliably greater than chance (i.e., 0.50).

Items recorded during the first iteration of the word list in the baseline and shadowing blocks were used in the AXB task (exceptions noted below). The first iteration of the shadowing block was presented directly after the exposure phase to ensure shadowers' familiarity with each native and accented speaker. Nevertheless, Babel *et al.* (2013) found no significant differences in alignment across consecutive shadowing blocks, so shadowing errors were replaced with tokens from the subsequent iteration. If shadowers made an error (uttered an incorrect target, e.g., "bail" instead of "wail," as judged by two coders) the first time a word was presented, a replacement was taken from the second iteration. Of the 163 shadowing errors, 51 (2.8% of shadowed utterances) were replaced with tokens from the second iteration and 112 tokens (7% of shadowed utterances) were eliminated from analysis because the shadower made errors on both shadowing presentations. Significantly fewer items were eliminated for native models (14/112) than Spanish-accented models (98/112), $X^2(1) = 63$, $p < 0.001$.

III. RESULTS

Mixed-effects regression models (MEMs) and logistic MEMs, where appropriate, were used to analyze the data. Maximal random effect structures justified by the sample were used, including random intercepts and slopes where appropriate (Barr *et al.*, 2013). In all models, predictors were deviation coded (-0.5 , 0.5). Stepwise model comparisons were used to justify the inclusion of fixed effects in final models (Baayen *et al.*, 2008; Jaeger, 2008). Analyses were conducted in the R environment (R development core team, 2013) using the AFEX (Singmann *et al.*, 2016) and LME4 (Bates *et al.*, 2014) packages. Multi-collinearity was assessed with Cramér's V for association between two nominal predictors (VCD; Meyer *et al.*, 2016), intra-class correlations for association between nominal and numeric predictors (ICC; Wolak, 2016), and kappa for association among numeric predictors

(LANGUAGER; Baayen, 2011). Unless otherwise noted, association among predictor variables was weak ($0 < \text{Cramér's } V < -0.0001 < \text{ICC} < 0.23$; $1.02 < \kappa < 1.461092$), indicating orthogonality. The study's primary focus concerns alignment as a function of native and non-native accent. However, because there were a small number of model talkers, analyses were also conducted by model talker to account for potential idiosyncratic talker effects.

A. Perceptual exposure

Accuracy (proportion correct items chosen from the 3×3 table) of responses in the perceptual exposure block was analyzed by quartiles. Trials 1–72 were considered quartile 1, 73–144 as quartile 2, etc. Quartiles contained equal numbers of items spoken by each model talker. Performance was near ceiling for both accent groups across the four quartiles (range 0.98–0.99). A mixed-effects model (MEM) containing accent ($\beta_{\text{Spanish}} = -0.581$, $\text{se} = 0.003$, $z = -2.57$, $p = 0.01$) as a fixed effect and shadower and word as random effects provided the best fit to the data [$X^2(1) = 6.865$, $p = 0.01$], reflecting lower accuracy for the Spanish ($\bar{x} = 0.98$) than for the English ($\bar{x} = 0.99$) utterances. Quartile ($\beta_{Q1} = 0.36$, $\beta_{Q2} = 0.43$, $\beta_{Q3} = 0.34$, $\beta_{Q4} = 0.28$; all $\text{se} < 2.5$, all $z > 0.55$, all $p < 0.14$) marginally improved the fit of the MEM to the data [$X^2(3) = 7.216$, $p = 0.07$], indicating that accuracy increased marginally over the course of exposure. The accent by quartile interaction was not significant, $X^2(3) = 1.393$, $p = 0.70$. Overall, performance was near ceiling for both Spanish and English utterances.

B. Research question 1A: Listener assessment of vocal alignment

Prior to analysis, responses for each trial of the AXB task were coded as 0 for BX responses (baseline utterance more similar to model) and 1 for AX responses (shadowed utterance more similar to model). This coding resulted in a proportion of perceived vocal alignment (see Fig. 1).

FIG. 1. Shows proportion of "shadowed more similar" AXB judgments plotted against model talker. Alignment to the native speakers, EF (English female) and EM (English male), and to the non-native speakers, SF (Spanish female) and SM (Spanish male), is significantly above chance (0.50), and the speaker groups differ significantly from each other. Error bars represent standard error of the mean and asterisks indicate significance at $p < 0.05$.

Logistic mixed-effects regression modeling (MEM) was used to analyze the data, as 0 and 1 bound the dependent measure. Association between model talker and accent was perfect (Cramér's $V = 1$) so they were tested in separate MEMs.

Based on the random effects only MEM (hereafter referred to as a "control MEM"), listeners' assessment of vocal alignment ($\bar{x} = 0.55$, $SD = 0.058$) was significantly greater than chance (0.50; $\beta = 0.207$, $z = 3.823$, $p < 0.005$) suggesting that, overall, listeners judged shadowers as having aligned to the model talkers. Talker's accent significantly improved the fit of the model, $X^2(1) = 24.015$, $p < 0.001$, indicating that model talker's native language influenced alignment. The β -estimate indicated that shadowers aligned more towards Spanish-accented model talkers ($\bar{x} = 0.57$) than to native talkers ($\bar{x} = 0.54$; $\beta_{\text{Spanish}} = 0.11$, $z = 4.899$, $p < 0.001$).

To examine potential differences in alignment across model talkers, a separate MEM was constructed with model talker as a fixed effect. Model talker improved the MEM's fit to the data, $X^2(3) = 41.85$, $p < 0.001$, indicating degree of alignment varied across model talkers. Follow-up MEMs indicated talkers EF and SF did not differ from each other ($\beta_{\text{SF}} = 0.04$, $z = 0.46$, $p = 0.64$), EM received less alignment than the other talkers ($\beta_{\text{EM}} = -0.23$, $z = -6.08$, $p < 0.001$), and SM received more alignment than the other talkers ($\beta_{\text{SM}} = 0.17$, $z = 4.13$, $p < 0.001$). These results show that, while there was overall more alignment to the Spanish than to the English talkers, alignment to model talkers was not uniform.

C. Research question 1B: Acoustic assessments of vocal alignment

Fundamental frequency (f_0), utterance duration, and vowel spectra ($F1$ and $F2$) were measured in PRAAT using custom scripts (Boersma and Weenink, 2013). Average F_0 was computed for each utterance. $F1$ and $F2$ were calculated from the midpoint of the vowel using LPC analysis. All three were Lobanov normalized to allow comparison across talkers while minimizing variance not due to idiolect or dialect (Adank *et al.*, 2004; Babel, 2012; Fabricius *et al.*, 2009; Lobanov, 1971). Inaccuracies (measures > 2 SD away from an individual's mean $F1$ or $F2$) were replaced with manual measurements. No data pruning procedures were used on F_0 or utterance duration measurements.

To index alignment, difference in distance scores (DID) were calculated for all acoustic measures by subtracting the absolute shadowed difference (shadowed-model) from the absolute baseline difference (baseline-model) for each word token (Pardo *et al.*, 2013). When calculated this way, a positive value indicates alignment (shadowed distance smaller than baseline distance). The magnitude of the DID score expresses extent of change, with larger values illustrating more change than smaller values. For normalized vowel formant measures, a Euclidean DID was calculated to capture vowel formants as relative points in a multi-dimensional space (Bradlow *et al.*, 1996). Mean F_0 , mean duration, and

mean vowel DID scores are shown in Figs. 2(A)–2(C), respectively.

Separate mixed-effects models (MEMs) were constructed to assess alignment on F_0 , duration, and vowel spectra, respectively. For each series of MEMs, a control MEM included shadower (with by-participants random slopes) and word as random effects. Significance was determined within an MEM and across models with MEM comparisons. Model talkers' accent (native or non-native) was the fixed effect of interest in all analyses. In the vowel analysis, the random intercept for word was perfectly associated with the fixed effect of vowel identity (Cramér's $V = 1$).

1. Fundamental frequency DID score

In order to determine if shadowers aligned to model talkers on normalized fundamental frequency (F_0), a control MEM was first constructed (as described in preceding

FIG. 2. Mean (A) F_0 DID score, (B) duration DID score, and (C) vowel DID score by model talker. In all panels, error bars represent standard error of the mean. Light bars indicate native English-speaking model talkers (ef, em) and dark bars indicate Spanish-accented model talkers (sf, sm). (A) shows a weak trend towards divergence on f_0 across both accent types. By contrast, (B) indicates that shadowers generally aligned their duration to model talkers, with a stronger effect for English model talkers than for Spanish. Finally, (C) shows small differences in Euclidean vowel DID score for all model talkers except for SM.

sections) with F0 DID score as the dependent measure. The control MEM's intercept was non-significant ($\beta = -9.6e-4$; $t = -0.02$, $p = 0.984$), indicating no significant overall alignment to F0. Accent did not improve model fit ($\beta = -0.04$; $t = -0.89$, $p = 0.38$; $X^2(1) = 5.6e-3$, $p = 0.94$, but model talker did, $X^2(3) = 63.3$, $p < 0.001$. *Post hoc* comparisons showed alignment to F0 differed significantly across talkers. SM ($\beta = 0.09$; $t = 2.1$, $p = 0.04$) received the most alignment on F0, alignment to EM ($\beta = 0.09$; $t = 2.26$, $p = 0.03$) and EF ($\beta = -0.03$; $t = -0.72$, $p = 0.48$) did not differ, and shadowers diverged from SF's F0 ($\beta = -1.6$; $t = -3.64$, $p = 0.001$). To summarize, this analysis revealed that F0 alignment varied by model talker but not by model talker's accent.

2. Duration DID score

The positive intercept in the control MEM indicated that alignment to duration was significantly greater than 0, $\beta = 32.22$; $t = 5.48$, $p < 0.001$. Adding accent ($\beta = -1.06$, $t = -0.29$, $p = 0.83$) did not improve the MEM's fit to the data [$X^2(1) = 0.026$, $p < 0.87$], indicating that shadowers did not differentially align toward native English and Spanish talkers' durations. Model talker improved the MEM's fit to the data, $X^2(3) = 1081.5$, $p < 0.001$. Follow-up MEMs indicated that there was greatest alignment to SM's ($\beta_{SM} = 26.5$, $t = 5.405$, $p < 0.001$) and EF's duration ($\beta_{EF} = 28.27$, $t = 5.77$, $p < 0.001$) and least alignment to SF's ($\beta_{SF} = -23.791$, $t = -4.910$, $p < 0.001$) and EM's duration ($\beta_{EM} = -31$, $t = -6.319$, $p < 0.001$). These results suggest that shadowers aligned to model talker's utterance duration and that the effect was better explained by model talker than by accent.

3. Vowel spectra DID score

A series of MEMs were constructed to examine the effect of accent on vocal alignment as assessed by Lobanov-normalized vowel DID score. Excluding vowels with few tokens (/aɪ/ /ɔ/ /ɛ/ /u/) did not alter the results, so analyses are reported on the full set. The control MEM indicated that vowel DID score did not significantly differ from zero ($\beta = 0.017$; $t = 1.16$, $p = 0.27$), suggesting that shadowers did not align to vowel spectra overall. Neither accent ($\beta_{Spanish} = 0.01$, $t = -0.49$, $p = 0.62$) nor model talker (all $\beta < 0.004$; all $t < 1.44$, all $p > 0.151$) significantly improved the MEM fit, $X^2 > 3.7$, $p > 0.3$. Next, vowel identity was entered into the MEM ($\beta_a = 0.183$, $\beta_{ai} = 0.085$, $\beta_o = -0.097$, $\beta_i = -0.014$, $\beta_e = 0.028$, $\beta_{ei} = 0.059$, $\beta_u = -0.024$, $\beta_u = -0.037$, $\beta_{ou} = -0.079$, $\beta_\Lambda = 0.012$). Vowel identity significantly improved model fit, $X^2(10) = 26.27$, $p < 0.001$, suggesting that degree of alignment observed in the vowel DID score differed as a function of vowel category. Although neither MEM comparisons nor changing the reference level revealed statistically reliable patterns of alignment by vowel, numerically the vowels /a/, /æ/, and /ɛ/ elicited greater alignment than other vowels.

D. Research question 2: Relationship between acoustic and perceptual indices of alignment

Although there was no evidence that shadowers aligned on F0 or overall on vowel spectra, these acoustic parameters may have contributed to the raters' perception of vocal alignment. Moreover, aligning to non-native speakers may produce distinct changes in shadowed productions that are differentially detectable by native English-speaking listeners. Analyses were conducted to examine the acoustic cues contributing to listeners' perceptions of alignment. For this set of MEMs, AXB ratings served as the dependent measure, and shadower, word, and AXB rater were entered as random effects. Accent and all acoustic measures of vocal alignment (F0 DID, duration DID, vowel spectra DID) served as fixed effects. Numeric fixed effects were *z*-transformed (i.e., centered) before being entered into the MEM to enable comparison across measures despite differences in the original scales. High collinearity was observed between F0 DID and vowel DID ($\kappa = 3.4e^{14}$).

The inclusion of accent ($\beta_{Spanish} = 0.11$, $z = 4.3$, $p < 0.001$), duration DID ($\beta_{DurationDID} = 0.18$, $z = 14.547$, $p < 0.001$), F0 DID score ($\beta_{F0DID} = 0.06$, $z = 5.396$, $p < 0.001$), and vowel spectra DID ($\beta_{vowelDID} = 0.036$, $z = 3.102$, $p = 0.002$) significantly improved MEM fit [$X^2(5) = 282$, $p < 0.001$], suggesting that the acoustic indices of alignment were related to listener perceptions of alignment. Furthermore, all acoustic measures contributed to the listeners' perceptions of vocal alignment. Figure 3 depicts these relationships.

After the relevant acoustic dimensions were identified, interactions between accent and the acoustic measures were examined for differences between talker groups. No significant interaction between duration DID score and accent ($\beta_{DurDID \times Spanish} = -0.04$, $z = -1.59$, $p = 0.11$) was found, $X^2(1) = 2.54$, $p = 0.111$. However, significant interactions were detected between F0 and accent ($\beta_{F0DID \times Spanish} = -0.09$, $z = -3.9$, $p < 0.001$), $X^2(3) = 18.025$, $p < 0.001$, and between vowel DID score and accent ($\beta_{vowelDID \times Spanish} = 0.06$, $z = 2.471$, $p = 0.01$), $X^2(3) = 11.932$, $p < 0.001$. The beta estimates indicate that alignment to vowel spectra of Spanish model talkers and to F0 of native model talkers led to greater perceptions of alignment by listeners.

A separate MEM included the three acoustic measures and model talker. Model talker significantly improved MEM fit [$X^2(3) = 41.85$, $p < 0.001$]. Follow-up analyses revealed interactions between model talker and acoustic measures of alignment. The talker by duration interaction significantly worsened MEM fit as determined by BIC and AIC, $X^2(3) = 46.116$, $p < 0.001$. The influence of alignment to F0 was equivalent for EM, EF, and SM (all $\beta < 0.05$, $z < 1.18$, $p > 0.12$), but that the effect of F0 DID was significantly less for SF ($\beta_{SF} = -0.12$, $z = -2.95$, $p = 0.003$). Talker comparison also indicated that alignment to vowel spectra was related to perceived alignment for EM and SF (both $\beta > 0.3$, $z > 2.9$, $p < 0.01$) and non-significantly for SM ($\beta = -0.02$, $z < -0.269$, $p = 0.78$). Divergence from EF's vowel spectra was associated with perceived alignment to EF ($\beta = -0.24$, $z = -2.69$, $p = 0.007$). For both F0 [$X^2(3) = 18.025$, $p < 0.001$] and vowel spectra [$X^2(3) = 11.932$, $p = 0.007$]

AXB Judgment by Acoustic Measures of Alignment

FIG. 3. The relationship between the acoustic and perceptual measures of alignment. For each panel, a single acoustic measure of alignment is plotted along the x axis and vocal alignment as perceived by the AXB raters is depicted on the y -axis. Points on the plot are individual shadowers, and the dotted lines indicate linear trend lines. While the linear trends across all three measures suggests that greater alignment to F0 (A), duration (B), and vowel spectra (C) is positively related to alignment on the AXB measure, only the relationships in (B) and (C) are statistically significant.

DID scores, including the interaction significantly improved MEM fit. These results indicate that greater alignment on acoustic measures is generally associated with higher perceptual ratings of alignment, though there are accent and talker-specific effects.

E. Research question 3: Baseline variability in speech production

Babel (2010) and Walker and Campbell-Kibler (2015) propose that vocal alignment is more likely to occur when differences between the baseline and target are large, but

only if the differences exist within the aligner's normal production repertoire. To test this claim for alignment to native- and Spanish-accented talkers, separate MEMs were constructed to assess the effect of variability in baseline acoustic properties of individual shadowers on vocal alignment. "Acoustic variability" was operationalized as standard deviation (SD) of F0, standard deviation (SD) of duration, and vowel dispersion of vowel spectra. These measures were calculated on a by-participant basis from baseline utterances only. In all MEMs, shadower and word were included as random effects. The relationship between each acoustic measure of variability and its corresponding DID score appears in Fig. 4. A multi-collinearity assessment indicated that word and vowel dispersion were highly associated, $ICC = 0.71$.

1. Fundamental frequency

As established by the analysis for research question 1B (Q1B), model talker significantly affected alignment to normalized F0. However, shadower's F0 SD ($\beta = -0.0004$, $t = 0.64$, $p = 0.52$) and the F0 SD by model talker interaction significantly worsened MEM fit as indexed by AIC and BIC, $X^2(4) = 95.36$, $p < 0.001$, suggesting that baseline flexibility in pitch did not affect degree of alignment to pitch.

2. Duration

Analysis from Q1B revealed significant alignment to duration that varied as a function of model talker. For this analysis, model talker was therefore included as a fixed effect. There was no main effect of duration SD ($\beta = 0.19$, $t = 0.407$, $p = 0.69$) or model talker by duration SD interactions (all $\beta < 0.97$, $t < 2.5$, $p > 0.02$). AIC and BIC indicated that the additional effects provided a significantly worse MEM than model talker alone, $X^2(3) = 22.17$, $p < 0.001$. The results indicate that variation in baseline duration did not predict degree of alignment to duration.

3. Vowel spectra

SD was used to define variation in F0 and duration, but *vowel dispersion* was used to index variability in vowel production. Vowel dispersion is typically characterized as a measure of speech precision but can arguably also measure a type of flexibility or variability. When vowel dispersion is high, it indicates that the speaker is using a greater portion of the available acoustic-phonetic space and therefore may be less, rather than more, restricted in the ability to produce a range of tokens. However, it should be noted that this measure does not capture variation or precision in the realization of *individual* vowels. As used here, vowel dispersion is a measure of the extent to which vowels differ from one another in the vowel space and hence, indexes a type of production variation that may be associated with productive flexibility.

Vowel dispersion was calculated by taking the Euclidean distance between every individual F1 and F2 value and the central F1 and F2 value for a participant (Bradlow *et al.*, 1996). The central value was calculated as the mean of each individual shadower's F1 and F2, and all

FIG. 4. The figures plot each measure of variability against its associated measure of vocal alignment. In all three panels, points on the plot represent means from individual shadowers, and the line is a linear regression line. All three regression lines are in the predicted direction (greater variability is associated with greater alignment on the corresponding acoustic measure). At a $p < 0.05$ threshold, the relationships in (A) and (B) are non-significant; and in (C), significant.

values were Lobanov-normalized prior to analysis. By this metric, a higher numerical value represents a more dispersed vowel space while a smaller value represents a more contracted vowel space. Because word was highly collinear ($ICC > 0.5$) with both vowel category and vowel dispersion, the random effect of word was excluded from the following analysis.

Baseline vowel dispersion ($\beta = 0.07$, $t = 11.92$; $p < 0.001$) significantly improved the fit over the MEM containing only vowel [$X^2(10) = 80.451$, $p < 0.001$]. The interaction between vowel dispersion and accent was non-significant, $X^2(2) = 1.479$, $p = 0.477$, suggesting that the effect of baseline vowel dispersion did not have a differential

impact on alignment to native and non-native model talkers. A significant model talker by vowel dispersion interaction indicated that higher vowel dispersion was associated with greater alignment for EF ($\beta_{EF \times Vdisp} = 0.07$, $t = 4.82$; $p < 0.001$), SM ($\beta_{SM \times Vdisp} = 0.07$, $t = 4.475$; $p < 0.001$), and SF ($\beta_{SF \times Vdisp} = 0.03$, $t = 2.11$; $p = 0.03$), but not for EM ($\beta = -0.08$, $t = -5.85$; $p < 0.001$). This result supports the proposal that greater degree of vowel dispersion or variability, as defined here, relates to greater alignment to vowel spectra.

IV. GENERAL DISCUSSION

The present experiment studied alignment to foreign- and native-accented speakers as a strategy for examining the interplay among cognitive, linguistic, and social mechanisms of vocal alignment. Several novel findings were observed. First, the native English speakers aligned to Spanish-accented model talkers as indexed by both perceptual and acoustic measures. Second, acoustic measures of alignment toward Spanish- and native-accented talkers were related to perceptual assessments, and that relationship differed by accent type. Third, greater baseline vowel dispersion was associated with greater vocal alignment on vowel spectra.

A. Vocal alignment to Spanish-accented speakers

Acoustic and perceptual assessments of alignment showed that shadowers aligned to both native and non-native speakers. Importantly, both measure types indicated that native talkers aligned to the Spanish models despite accent-ness, intelligibility, and attitude differences between the non-native and native speakers. Indeed, although shadowers did not demonstrate differential alignment on any of the measured acoustic parameters, listener assessments of alignment showed that native English shadowers aligned *more* to the non-native English targets. The observed pattern of alignment corresponded to differences in model talker intelligibility and accentedness rather than to attitudinal differences toward accented talkers, consistent with perceptual-motor and action understanding accounts. The current findings imply that in the context of a minimally social shadowing paradigm, alignment behaviors may primarily be driven by putative links between speech perception and production that allow the perception of unfamiliar and/or distinctive talker- and group-specific variation to be reflected in speech production (Fowler and Galantucci, 2005; Goldinger, 1998; Pickering and Garrod, 2004, 2013). Vocal alignment may be used to compensate for the lower intelligibility of the non-native speakers and to help listeners recover linguistic content (Adank *et al.*, 2010). Alternatively, the non-standard properties of Spanish-accented English may have violated listeners' expectations and captured more attention, thereby increasing vocal alignment (Babel *et al.*, 2014; Jaeger and Weatherholtz, 2016). Although the current paradigm cannot disambiguate between these two possibilities, both highlight the relative importance of perceptual properties of speech and the extent to which sensitivity to these properties influences the production of spoken language.

The few studies that have examined vocal alignment to non-native speakers have inconsistently demonstrated alignment (Kim *et al.*, 2011; Kim, 2011, 2012). Kim *et al.* (2011) found that a conversational setting yielded alignment in matched-language dyads, but not in mixed-language dyads, putting the current results seemingly at odds. However, Kim *et al.* observed that the native English speakers in their study may have engaged in a conscious or effortful decision, based either on social attitudes or linguistic goals (e.g., to be more globally intelligible), not to align to the accented talkers. Unlike Kim *et al.*, the present study used a task that minimized the relevance of sociolinguistic communicative strategies and goals. When explicit reasons to adopt a non-alignment communicative strategy were absent, native English speakers vocally aligned to non-native speakers. This difference may account for the disparity across studies, and further offer a unique contribution to the understanding of vocal alignment. When considered together, the findings support the role of both perceptual-motor and social-motivational components in vocal alignment. Crucially, the current experimental paradigm implies that perceptual-motor couplings may act in an automatic way to contribute to alignment behavior, while the study of Kim *et al.* study suggests a potential route for effortful or conscious control.

Social-motivational accounts would predict that the negative attributions associated with non-native speakers lead to less vocal alignment (Giles *et al.*, 1991). However, the current findings demonstrate generally more rather than less alignment to the accented talkers and certainly the characteristics of the non-native speakers did not prohibit or block alignment. As research continues, mechanistic accounts of vocal alignment must consider how the low-level perceptual-production mechanism observed in the current study interacts with social-motivational information. To this end, accounts of behavioral alignment (e.g., alignment to gesture, facial expression) may provide a useful model for alignment in the speech domain (Chartrand and Bargh, 1999; Louwerse *et al.*, 2012). Behavioral alignment may be largely automatic, with a potential secondary mechanism regulating alignment based on situational factors (Van Baaren *et al.*, 2009). Such a dual mechanism may be operating in the speech domain as well.

Another factor that may have contributed to the observation of alignment to non-native English speakers in the current study is the inclusion of perceptual exposure trials. In addition to generally acquainting shadowers with the idiosyncrasies of the four model talkers, the exposure may have also increased perceptual fluency for the non-native speakers. “Perceptual fluency” is the ease with which a stimulus can be processed, identified, or understood (Reber *et al.*, 2004). Dragojevic and Giles (2016) find that speech that lacks perceptual fluency (i.e., speech embedded in white noise, non-native speech) may prompt more negative attitudes toward the speaker. Perceptual exposure in the current study may have made the non-native speech easier or more fluent to process, potentially lessening negative attributions made of the particular model talkers. Perceptual exposure may also have highlighted the acoustic differences between native and non-native speech productions, encouraging the shadowers

to map unfamiliar phonological forms onto their own representational systems (Adank *et al.*, 2010; Kappes *et al.*, 2009; Tzeng *et al.*, 2016). The specific contribution of talker familiarity to vocal alignment remains for future research.

B. Relationship between acoustic and perceptual measures

Although many studies have explored the acoustic correlates of vocal alignment (e.g., Delvaux and Soquet, 2007; Miller *et al.*, 2013; Yu *et al.*, 2013), relatively few studies have incorporated both acoustic and perceptual measures (Babel *et al.*, 2013; Kim, 2012; Pardo *et al.*, 2013; Shockley *et al.*, 2004; Walker and Campbell-Kibler, 2015), and even fewer have examined the relationship among them (Abel and Babel, 2017; Pardo *et al.*, 2013).

In the present analyses, shadowers who exhibited greater alignment to model talkers’ fundamental frequency, duration and vowel spectra (as indicated by larger DID scores) received higher perceptual ratings of alignment. For duration, there was no effect of model talker or accent group, meaning that this was a global target of alignment and not specific to a particular accent group or individual. However, both F0 DID and vowel DID interacted with model talkers’ accent such that F0 predicted listeners’ judgments of alignment to native speakers and vowel DID predicted listeners’ judgments of alignment to Spanish-accented speech. This suggests that some dimensions of alignment may be specific to a particular talker or accent group, based on large, salient differences in vocal style associated with that talker or group (Delvaux and Soquet, 2007; Mitterer and Müssler, 2013). The current paper thus adds to a small body of work that identifies acoustic dimensions of alignment and relates them to how listeners in an AXB task may form judgments of alignment.

C. Individual differences in variability and vocal alignment

That shadowers in the current study varied extensively in degree of vocal alignment (AXB alignment range = 0.45–0.67) highlights the importance of understanding individual differences in shadower characteristics. For example, some evidence suggests that males and females align to different extents and on different dimensions (Biliouss and Krauss, 1988; Namy *et al.*, 2002; Pardo, 2006). Other recent studies have found that individual characteristics such as an individual’s rejection sensitivity (Aguilar *et al.*, 2016) and personality traits (openness and agreeableness; Yu *et al.*, 2013) may affect individuals’ tendency to align. The current experiment offers some insights regarding the potential contribution of shadowers’ articulatory flexibility to individual vocal alignment behavior.

Walker and Campbell-Kibler (2015) propose that vocal alignment across dialects of a language (e.g., New Zealand English and US Midland English) may be most likely when cross-dialect differences are large and the variants to be shadowed exist within a talker’s baseline phonetic repertoire [see Babel (2010) and Kim *et al.* (2011) for a similar argument]. The current study explored an extension of this

proposal by assessing variation in individual speakers' productions and the relationship between that variation and the extent of vocal alignment to native and non-native speakers.

In the current study, baseline variation in speech production was assessed with standard deviations (for F0 and duration) and vowel dispersion (for vowel spectra) on a by-subject basis. As noted above, the vowel dispersion measure assessed "variation" in the sense that larger vowel spaces might be associated with greater productive variation. Indeed, those shadowers who exhibited greater baseline variation in vowel spectra appeared to exhibit greater vocal alignment as indexed by the corresponding vowel DID measure. Individuals may vary in basic articulatory variability (i.e., proportion of the articulatory vowel space used), and their own flexibility may predict their ability to match the speaking characteristics of others. We speculate that this flexibility may be related to social factors that influence alignment, such as personality or gender (e.g., Yu *et al.*, 2013). For example, gender differences that some researchers find in vocal alignment research may be partially explained in terms of articulatory flexibility rather than strictly in terms of differences in socialization (though the two may covary; Biliouss and Krauss, 1988; Namy *et al.*, 2002). Males and females tend to differ in their articulatory precision (Bradlow *et al.*, 1996), and cross-culturally, the differences between male and female speech are greater than would be expected from anatomy alone (Johnson, 2006). If females tend to exhibit greater articulatory flexibility than males, it may lead to the observation that females align more than males. The present study's sample was not selected to explore gender differences, so it remains for future work to test this possibility. The current findings do, however, suggest articulatory flexibility may contribute to the individual differences in alignment behavior that are observed across studies.

D. Limitations and future directions

The number of model talkers, while consistent with the number typically examined in shadowing studies, captures only a small amount of possible talker variation. With only two talkers representing native and non-native speech, it is difficult to claim that alignment patterns will be the same for all other English and Spanish-accented talkers. Indeed, although shadowers clearly aligned to the non-native speakers, the extent of alignment also varied as a function of individual talker, consistent with previous research (Namy *et al.*, 2002; Pardo *et al.*, 2013; Pardo *et al.*, 2017). Nevertheless, although the current study does demonstrate alignment *despite* differences in native language background, future research should continue to consider the contributions of intelligibility and attitudes toward individual model talkers. Relatedly, the effects of gender on both vocal alignment and the detection of vocal alignment are unresolved and the large number of female participants, both shadowers and AXB raters, may have influenced the results. Because the study sample was unbalanced with respect to sex, it was not possible to adequately address this factor.

Another consideration is the degree of similarity between the model talkers and the shadowers. The Spanish-accented talkers in the current study were selected because their speech characteristics are distinct from the native

talkers. Thus, the AXB listeners may have had different thresholds for judging similarity between the native model talkers and shadowers and between the non-native model talkers and shadowers. Listeners in the AXB task were not asked how they were making their judgments; therefore, their threshold for what can be considered similar is not known. However, alignment on three acoustic DID measures predicted higher perceived similarity in the AXB task. This suggests that, while similarity thresholds for making AXB ratings are not known, listeners in the AXB task were sensitive to similarity in the acoustic space.

The current study may also have underestimated the amount of alignment to the non-native speakers. When shadowers produced tokens such as [kin] in response to a Spanish model talker, it was unclear whether the shadower was aligning to the Spanish vowel inventory (/i/ → /i/) or because the word "kin" was instead misperceived as "keen." To avoid the issue of adjudicating the cause (vocal alignment or genuine error) of such a production, alignment was assessed by within-category variation rather than between-category substitution (e.g., kin → kin). However, this procedure may have excluded instances of genuine alignment.

V. CONCLUSION

The current study provides an examination of the underlying mechanisms responsible for speech alignment. More vocal alignment to Spanish-accented than to native speakers was observed, despite lower ratings for non-native speakers on measures of status, suggesting that absent other constraints on the alignment process, vocal alignment is driven in part by shared or linked representations between speech perception and speech production. Furthermore, although the relationship between acoustic measures and holistic, listener-based representations of alignment do not correspond perfectly, it is possible to identify acoustic dimensions of alignment and to assess how those dimensions differ as a function of the accentedness of the talker. Finally, an individual's natural range of vowel productions may constrain the extent to which he or she vocally aligns. This finding represents one of the first examinations of the role of habitual speech variation on an individual's propensity to align.

ACKNOWLEDGMENTS

Portions of this work were presented at the 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, HI, 2016 and at the 55th Annual Meeting of the Psychonomic Society, Long Beach, CA, November 2014. The authors would like to thank Valentin Lazar for his help in programming the experimental task, and Sana Bhojwani, Soo Mhin Park, and Sophie Cohn for their assistance preparing stimuli and data for analysis. Thanks also to Laura Namy, Harold Gouzoules, and Christina Tzeng for their feedback on earlier versions of this manuscript. This research was supported in part by Research Grant No. R01 DC 008108 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

APPENDIX A: WORD LIST

List of words used in both visual and auditory presentation formats.

ban	hag	pat
bead	hash	pawn
beak	hick	pet
bean	hid	pup
bud	hoot	rat
bug	hum	rhyme
bum	kin	rim
bun	kit	rum
chat	knob	rut
cheer	lace	sane
chore	lad	sill
cod	lame	soak
comb	lice	tan
con	mace	teat
cot	main	toot
dame	mall	wad
den	mat	wade
doom	mid	wail
dune	mitt	wed
fade	moan	weed
fin	moat	white
goat	mole	whore
gut	mum	wick
hack	pad	wrong

APPENDIX B: ATTITUDES RATING TASK

Sixteen independent listeners provided ratings on 21 attribute pairs for each model talker to ensure that the auditory stimuli used in the current experiment elicited the

TABLE II. The 21 attribute pairs used in the attitudes rating task, with corresponding factor loadings.

Attribute			Components	
			1	2
Disorganized	-	Well-organized	0.628	0.251
Dumb	-	Smart	0.715	0.312
Incompetent	-	Competent	0.767	0.139
Lacks confidence	-	Self-confident	0.490	-0.089
Poor	-	Rich	0.800	-0.227
Poor English	-	Good English	0.700	-0.081
Poorly Educated	-	Well-educated	0.894	0.003
Unimportant person	-	Important person	0.687	0.194
Unintelligent	-	Intelligent	0.810	0.271
Not sociable	-	Sociable	0.288	0.263
Arrogant	-	Humble	-0.498	0.564
Crude	-	Polite	-0.037	0.782
Dishonest	-	Honest	0.235	0.618
Dislike	-	Like	0.073	0.633
Gloomy	-	Cheerful	0.262	0.430
Mean	-	Nice	0.082	0.849
Rude	-	Cordial	0.110	0.785
Unfriendly	-	Friendly	-0.06	0.807
Unreliable	-	Reliable	0.432	0.525
Unsympathetic	-	Sympathetic	-0.235	0.722
Untrustworthy	-	Trustworthy	0.395	0.619

predicted differences in social attitudes. Ratings were collapsed across model talker to yield a mean for each accent group prior to analysis. Because assessing pairwise differences in ratings of 21 attributes would have led to a substantial multiple comparisons problem (i.e., at least one type I error guaranteed at $\alpha = 0.05$), attribute pairs were reduced into two categories based on a factor analysis with Varimax rotation in SPSS (IBM Corp., 2013).

A two-factor solution accounted for 53% of the variance in the data. The fit of a two-factor solution indicates that listeners evaluated model talkers along a minimum of two dimensions. Once the two factors were identified, rating scores were collapsed across attributes that loaded significantly (>0.40) onto each (see Table II). The first factor included attributes thought to reflect competency or status, including *intelligent*, *well-educated*, *rich*, etc. (9 attributes; Preston, 1999). *Friendly*, *polite*, *nice*, and other traits reflecting sociality or solidarity tended to load onto the second factor (11 attributes; Heaton and Nygaard, 2011; Preston, 1999).

¹See supplementary material at <https://doi.org/10.1121/1.5038567> for summary statistics of the model talkers' acoustic properties (Fig. S1 and Table S1).

- Abel, J., and Babel, M. (2017). "Cognitive load reduces perceived linguistic convergence between dyads," *Lang. Speech* **60**(3), 479–502.
- Abrego-Collier, C., Grove, J., Sonderegger, M., and Yu, A. C. (2011). "Effects of speaker evaluation on phonetic convergence," in *Proceedings of the 17th International Congress of the Phonetic Sciences*, pp. 192–195.
- Adank, P., Hagoort, P., and Bekkering, H. (2010). "Imitation improves language comprehension," *Psychol. Sci.* **21**(12), 1903–1909.
- Adank, P., Smits, R., and Van Hout, R. (2004). "A comparison of vowel normalization procedures for language variation research," *J. Acoust. Soc. Am.* **116**(5), 3099–3107.
- Aguilar, L., Downey, G., Krauss, R., Pardo, J., Lane, S., and Bolger, N. (2016). "A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matters," *J. Pers.* **84**(2), 165–177.
- Atagi, E., and Bent, T. (2015). "Relationship between listeners' nonnative speech recognition and categorization abilities," *J. Acoust. Soc. Am.* **137**(1), EL44–EL50.
- Atagi, E., and Bent, T. (2017). "Nonnative accent discrimination with words and sentences," *Phonetica* **74**(3), 173–191.
- Babel, M. (2010). "Dialect divergence and convergence in New Zealand English," *Lang. Soc.* **39**, 437–456.
- Babel, M. (2012). "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," *J. Phon.* **40**(1), 177–189.
- Babel, M., and Bulatov, D. (2012). "The role of fundamental frequency in phonetic accommodation," *Lang. Speech* **55**(2), 231–248.
- Babel, M., McAuliffe, M., and Haber, G. (2013). "Can mergers-in-progress be unmerged in speech accommodation?," *Front. Psychol.* **4**, 653.
- Babel, M., McGuire, G., Walters, S., and Nicholls, A. (2014). "Novelty and social preference in phonetic accommodation," *Lab. Phonol.* **5**(1), 123–150.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *J. Mem. Lang.* **68**, 255–278.
- Baayen, R. H. (2011). "languageR: Data sets and functions with 'Analyzing Linguistic Data: A practical introduction to statistics' R package version 1.4," <http://CRAN.R-project.org/package=languageR> (Last viewed 5 February 2018).
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). "Mixed-effects modeling with crossed random effects for subjects and items," *J. Mem. Lang.* **59**, 390–412.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). "lme4: Linear mixed-effects models using Eigen and S4," R package version 1.1-4, <https://github.com/lme4/lme4/> (Last viewed 5 February 2018).

- Bilious, F. R., and Krauss, R. M. (1988). "Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads," *Lang. Commun.* 8(3/4), 183–194.
- Boersma, P., and Weenink, D. (2013). "Praat: Doing phonetics by computer" [computer program], version 5.3.51, <http://www.praat.org/> (Last viewed 2 June 2013).
- Bourhis, R. Y., and Giles, H. (1977). "The language of intergroup distinctiveness," in *Language, Ethnicity and Intergroup Relations*, edited by H. Giles (Academic, London), pp. 119–135.
- Bradlow, A. R. (1995). "A comparative acoustic study of English and Spanish vowels," *J. Acoust. Soc. Am.* 97(3), 1916–1924.
- Bradlow, A. R., and Bent, T. (2008). "Perceptual adaptation to non-native speech," *Cognition* 106(2), 707–729.
- Bradlow, A. R., and Pisoni, D. B. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," *J. Acoust. Soc. Am.* 106, 2074–2085.
- Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech. I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* 20(3), 255–272.
- Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., and Shearman, S. M. (2002). "Attitudinal and affective response toward accented English," *Lang. Commun.* 22, 171–185.
- Brysbaert, M., and New, B. (2009). "Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English," *Behav. Res. Meth.* 41(4), 977–990.
- Chartrand, T. L., and Bargh, J. A. (1999). "The chameleon effect: The perception-behavior link and social interaction," *J. Pers. Soc. Psychol.* 76(6), 893–910.
- Delvaux, V., and Soquet, A. (2007). "The influence of ambient speech on adult speech productions through unintentional imitation," *Phonetica* 64, 145–173.
- Dias, J. W., and Rosenblum, L. D. (2016). "Visibility of speech articulation enhances auditory phonetic convergence," *Atten. Percept. Psychophys.* 78(1), 317–333.
- Dragojevic, M., and Giles, H. (2016). "I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process," *Hum. Commun. Res.* 42, 396–420.
- Dragojevic, M., Giles, H., and Gasiorek, J. (2015). "Communication accommodation theory," in *The International Encyclopedia of Interpersonal Communication*, edited by C. R. Berger and M. E. Roloff.
- Fabricius, A. H., Watt, D., and Johnson, D. E. (2009). "A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics," *Lang. Var. Change* 21, 413–425.
- Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (2003). "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks," *J. Mem. Lang.* 49(3), 396–413.
- Fowler, C. A., and Galantucci, B. (2005). "The relation of speech perception and speech production," *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Publishing Ltd., Malden, MA), pp. 632–652.
- Garrod, S., and Pickering, M. J. (2013). "Dialogue: Interactive alignment and its implications for language learning and language change," in *The Language Phenomenon* (Springer, Berlin), pp. 47–64.
- Giles, H. (1973). "Accent mobility: A model and some data," *Anthropol. Ling.* 15(2), 87–105, available at <http://www.jstor.org/stable/30029508>.
- Giles, H., Coupland, N., and Coupland, J. (1991). "Accommodation theory: Communication, context, and consequence," in *Studies in Emotion and Social Interaction. Contexts of Accommodation: Developments in Applied Sociolinguistics* (Cambridge University Press, New York), pp. 1–68.
- Giles, H., and Ogay, T. (2007). "Communication accommodation theory," in *Explaining Communication: Contemporary Theories and Exemplars*, edited by B. B. Whaley and W. Samter (Lawrence Erlbaum, Mahwah, NJ), pp. 293–310.
- Gluszek, A., and Dovidio, J. F. (2010). "The way they speak: A social psychological perspective on the stigma of nonnative accents in communication," *Pers. Soc. Psychol. Rev.* 14(2), 214–237.
- Goldinger, S. D. (1998). "Echoes of echoes? An episodic theory of lexical access," *Psychol. Rev.* 105(2), 251–279.
- Goldinger, S. D., Pisoni, D. B., and Logan, J. S. (1991). "On the nature of talker variability effects on recall of spoken word lists," *J. Exp. Psychol. Learn. Mem. Cogn.* 17(1), 152–162.
- Gregory, D., and Webster, S. (1996). "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status predictions," *J. Pers. Soc. Psychol.* 70, 1231–1240.
- Heaton, H., and Nygaard, L. C. (2011). "Charm or harm: Effect of passage content on listener attitudes toward American English accents," *J. Lang. Soc. Psychol.* 30(2), 202–211.
- IBM Corporation (2013). "IBM SPSS statistics for Windows," version 22.0 (IBM Armonk, NY).
- Jaeger, T. F. (2008). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," *J. Mem. Lang.* 59, 434–446.
- Jaeger, T. F., and Weatherholtz, K. (2016). "What the heck is salience? How predictive language processing contributes to sociolinguistic perception," *Front. Psychol.* 7, 1115.
- Johnson, K. (2006). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," *J. Phonetics* 34, 485–499.
- Kappes, J., Baumgärtner, A., Peschke, C., and Ziegler, W. (2009). "Unintended imitation in non-word repetition," *Brain Lang.* 111, 140–151.
- Kim, M. (2011). "Phonetic convergence after perceptual exposure to native and nonnative speech: Preliminary findings based on fine-grained acoustic-phonetic measurement," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 1074–1077.
- Kim, M. (2012). "Phonetic accommodation after auditory exposure to native and nonnative speech," Ph.D. dissertation, Northwestern University, Evanston, IL.
- Kim, M., Horton, W. S., and Bradlow, A. R. (2011). "Phonetic convergence in spontaneous conversations as a function of interlocutor language distance," *Lab. Phonol.* 2(1), 125–156.
- Kinzler, K. D., Shutts, K., DeJesus, J., and Spelke, E. S. (2009). "Accent trumps race in guiding children's social preferences," *Soc. Cognition* 27(4), 623–634.
- Lieberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," *Cognition* 21, 1–36.
- Lobanov, B. (1971). "Classification of Russian vowels spoken by different speakers," *J. Acoust. Soc. Am.* 49(2), 606–608.
- Louwerse, M. M., Dale, R., Bard, E. G., and Jeuniaux, P. (2012). "Behavior matching in multimodal communication is synchronized," *Cognitive Sci.* 36, 1–24.
- McKernan, D. J., and Hamayan, E. V. (1984). "Speech norms and attitudes toward outgroup members: A test of a model in a bicultural context," *J. Lang. Soc. Psychol.* 3(1), 21–38.
- Meyer, D., Zeileis, A., Hornik, K., Gerber, F., and Friendly, M. (2016). "vcd: Visualizing categorical data," R package version 1.4-3, <https://CRAN.R-project.org/package=vcd> (Last viewed 29 November 2017).
- Miller, R. M., Sanchez, K., and Rosenblum, L. D. (2013). "Is speech alignment to talkers or tasks?," *Atten. Percept. Psychophys.* 75(8), 1817–1826.
- Mitterer, H., and Müsseler, J. (2013). "Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech," *Atten. Percept. Psychophys.* 75(3), 557–575.
- Munro, M. J., and Derwing, T. M. (1995). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," *Lang. Learn.* 45(1), 73–97.
- Munro, M. J., Derwing, T. M., and Flege, J. E. (1999). "Canadians in Alabama: A perceptual study of dialect acquisition in adults," *J. Phon.* 27(4), 385–403.
- Namy, L. L., Nygaard, L. C., and Sauerteig, D. (2002). "Gender differences in vocal accommodation: The role of perception," *J. Lang. Soc. Psychol.* 21(4), 422–432.
- Nelson, L. R., Signorella, M. L., and Botti, K. G. (2016). "Accent, gender, and perceived competence," *Hispanic J. Behav. Sci.* 38(2), 166–185.
- Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific learning in speech perception," *Percept. Psychophys.* 60(3), 355–376.
- Pardo, J. (2013). "Reconciling diverse findings in studies of phonetic convergence," *Proc. Mtgs. Acoust.* 19, 060140.
- Pardo, J., Gash, H., Urmanche, A., Decker, A., Francis, K., Wiener, J., and Parker, S. (2014). "Effects of talker sex on phonetic convergence in shadowed speech," *J. Acoust. Soc. Am.* 135(4), 2420–2420.
- Pardo, J. S. (2006). "On phonetic convergence during conversational interaction," *J. Acoust. Soc. Am.* 119(4), 2382–2393.
- Pardo, J. S., Jay, I. C., and Krauss, R. M. (2010). "Conversational role influences speech imitation," *Atten. Percept. Psychophys.* 72(8), 2254–2264.
- Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., and Lewandowski, E. (2013). "Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures," *J. Mem. Lang.* 69(3), 183–195.
- Pardo, J. S., Urmanche, A., Wilman, S., and Wiener, J. (2017). "Phonetic convergence across multiple measures and model talkers," *Atten., Percept., Psycho.* 79(2), 637–659.

- Park-Taylor, J., Ng, V., Ventura, A. B., Kang, A. E., Morris, C. R., Gilbert, T., Devika, S., and Androsiglio, R. A. (2008). "What it means to be and feel like a 'true' American: Perceptions and experiences of second-generation Americans," *Cultur. Divers. Ethnic Minor. Psychol.* **14**, 128–137.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Pickering, M. J., and Garrod, S. (2004). "Toward a mechanistic psychology of dialogue," *Behav. Brain Sci.* **27**(2), 169–190.
- Pickering, M. J., and Garrod, S. (2013). "An integrated theory of language production and comprehension," *Behav. Brain Sci.* **36**(04), 329–347.
- Prinz, W. (1997). "Perception and action planning," *Eur. J. Cogn. Psychol.* **9**(2), 129–154.
- Preston, D. (1999). "A language attitude approach to the perception of regional variety," in *Handbook of Perceptual Dialectology*, edited by D. Preston (John Benjamins, Amsterdam, Netherlands), pp. 359–373.
- R development core team (2013). "R: A language and environment for statistical computing" (R Foundation for Statistical Computing, Vienna), <http://www.R-project.org> (Last viewed 3 March 2018).
- Reber, R., Schwarz, N., and Winkielman, P. (2004). "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?," *Pers. Soc. Psychol. Rev.* **8**(4), 364–382.
- Ryan, E. B. (1983). "Social psychological mechanisms underlying native speaker evaluations of non-native speech," *Stud. Second Lang. Acquis.* **5**, 148–159.
- Sancier, M. L., and Fowler, C. A. (1997). "Gestural drift in a bilingual speaker of Brazilian Portuguese and English," *J. Phonetics* **25**, 421–436.
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime: User's Guide* (Psychology Software Incorporated).
- Shockley, K., Sabadini, L., and Fowler, C. A. (2004). "Imitation in shadowing words," *Percept. Psychophys.* **66**(3), 422–429.
- Šidák, Z. (1967). "Rectangular confidence regions for the means of multivariate normal distributions," *J. Am. Stat. Assoc.* **62**(318), 626–633.
- Sidas, S. K., Alexander, J. E., and Nygaard, L. C. (2009). "Perceptual learning of systematic variation in Spanish-accented speech," *J. Acoust. Soc. Am.* **125**, 3306–3316.
- Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2016). "Afex: Analysis of factorial experiments," <https://CRAN.R-project.org/package=afex> (Last viewed 5 February 2018).
- Staum-Casasanto, L., Jasmin, K., and Casasanto, D. (2010). "Virtually accommodating: Speech rate accommodation to a virtual interlocutor," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, edited by S. Ohlsson and R. Catrambone (Cognitive Science Society, Austin, TX), pp. 127–132.
- Sumner, M. (2011). "The role of variation in the perception of accented speech," *Cognition* **119**(1), 131–136.
- Tzeng, C. Y., Alexander, J. E. D., Sidas, S. K., and Nygaard, L. C. (2016). "The role of training structure in perceptual learning of accented speech," *J. Exp. Psychol. Human* **42**(11), 1793–1805.
- Van Baaren, R., Janssen, L., Chartrand, T. L., and Dijksterhuis, A. (2009). "Where is the love? The social aspects of mimicry," *Philos. Trans. R. Soc. B.* **364**, 2381–2389.
- Walker, A., and Campbell-Kibler, K. (2015). "Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task," *Front. Psychol.* **6**, 1–18.
- Wolak, M. (2016). "ICC: Facilitating estimation of the intraclass correlation coefficient," r package version 2.3.0, <http://github.com/matthewwolak/ICC> (Last viewed 29 November 2017).
- Yu, A. C., Abrego-Collier, C., and Sonderegger, M. (2013). "Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and 'autistic' traits," *PloS One* **8**(9), e74746.