Xjenza Online - Journal of The Malta Chamber of Scientists
www.xjenza.org
Doi: http://dx.medra.org/10.7423/XJENZA.2014.1.02

Research Article

Agreeing to disagree: Constant non-alignment of speech gestures in dialogue

Holger Mitterer[1]
[1]Faculty of Media and Knowledge Sciences, Department of Cognitive Science, University of Malta

Abstract. Numerous studies suggest that interlocutors in a dialogue align with each other in terms of their articulatory gestures. It is often suggested that this, first, is the consequence of an automatic tendency for imitation and, second, this fosters mutual understanding. Making use of online archives of media, it was tested whether alignment is hence inevitable. The focus was on the pronunciation of the German word *ist* (Engl., 'is'). The standard pronunciation is [ɪst], but speakers with a Swabian accent produce [ɪʃt], acoustically reflected in the fricative spectra. We measured the spectra of fricatives in *ist* from interviewers while interviewing either a prominent German politician using the Swabian variant or an interviewee using the standard variant. Results showed neither an overall influence of the interviewees' pronunciation on the fricative realization by the interviewer nor a tendency to align over time for interviewer-interviewee pairs with different pronunciations. This shows that phonetic alignment in conversation is a more complex process than most current theories seem to suggest. Moreover, failure to align may not impede mutual understanding.

1 Introduction

While speech production and speech perception are typically studied as separate phenomena, the speech we hear influences how we speak. A seminal study by (Harrington et al., 2000) showed that speech production remains flexible over the lifespan and adapts to the ambient speech see also (Sancier and Fowler, 1997). Numerous studies have shown similar effects in laboratory settings, with imitation occurring over the time span of a conversation (Pardo, 2006) and in laboratory tasks such as shadowing (Fowler et al., 2003; Mitterer and Ernestus, 2008; Shockley et al., 2004). In a broader framework, (Pickering and Garrod, 2004) suggested that such alignment effects occur on many levels in interactive dialogue and that they underlie the ease with which interlocutors achieve parity, see also (Miller et al., 2010).

Phonetic alignment has been found with both phonetic measurements and native listeners' judgments. (Fowler et al., 2003), for instance, measured the voice onset time (VOT) of stop-vowel syllables recorded during a shadowing task. The to-be-shadowed stimuli had a long or a short VOT, and this affected the VOT in the shadowing responses. (Brouwer et al., 2010) measured the duration and the degree of segment reduction for shadowed utterances in response to more or less casual speech. They found that shadowing responses to strongly casual speech were shorter and contained more segment reductions than responses to more formal speech. Another frequently employed method to test for phonetic alignment is a perceptual matching task (Goldinger, 1998). In this task, participants hear utterances from a given speaker recorded in a baseline condition or recorded during or after conversation with a reference speaker. Participants typically judge the utterances recorded during or after the conversation as more similar to utterances from the reference speaker than the baseline recordings (Pardo, 2006), indicating phonetic alignment.

Maybe due to the practice of "null-hypothesis significance testing", in which the absence of alignment is nothing more than an uninformative null-effect, the emerging picture seems to be that one adapts to all aspects of an interlocutor's speech. As (Miller et al., 2010) (p. 1615) state "[i]n summary, speech alignment occurs

Correspondence to: H. Mitterer (holger.mitterer@um.edu.mt)

on both phonetic and extraphonetic levels". Intuitively, however, it seems unlikely that one would align with all aspects with an interlocutor, independent of the difference between the interlocutor's and one's own pattern. An obvious example here is a foreign accent. It seems unlikely that a native speaker would incorporate aspects of a foreign accent in one's own speech.

Indeed, recent studies with both acoustic and perceptual measures of alignment often report perceptual alignment but no acoustic alignment (Pardo et al., 2013). One possible reason for this may be that there were no active manipulation of phonetic differences between stimuli and responses, that is, speakers and shadowers were from the same sociolinguistic background and there was no a-priori defined phonetic difference between the two groups. It may be necessary to start with such differences. Indeed, another study (Nguyen et al., 2012) investigated a dialect difference between different versions of French and found evidence for convergence. Importantly, this study showed that Northern French speakers, considered to be the standard, adapt to non-standard Southern French speakers. This raises the question whether convergence in terms of speech gestures used for a given segment is inevitable. We identified a source of naturalistic data to test this in German fricative productions. German contrasts an alveolar fricative [s] with a post-alveolar fricative [ʃ] (similar to the English minimal pair 'sea'-'she'). In fricative-stop clusters, however, the contrast is neutralized, and Standard German uses only [ʃt] in onset position (e.g., 'Stein' / ʃteɪn/, Engl. 'stone') and only /st/ in coda position (e.g., 'Faust' /faust/). Swabian speakers from the Southwest of Germany, however, use [ʃt] in both cases. The difference between the two fricatives is captured well by a measurement of the spectral center of gravity. For /s/, the tongue is close to the teeth, and the resulting small space gives rise to high-frequencies, while for /ʃ/, the tongue is further back, leaving more space between tongue tip and teeth, given rise to lower frequencies: Just as for musical instruments, more space means lower frequencies. The center of gravity is something like the "average" frequency of the signal and is hence lower for /ʃ/ than for /s/. Given this acoustic correlate, it is straightforward to measure whether there is alignment or not.

If there is alignment when a Standard German speaker interacts a Swabian speaker, either the standard speaker, who regular produces [st] coda clusters, should produce fricatives in these clusters with an increasingly lower spectral center of gravity—becoming more [ʃ] like, or the Swabian speaker, who regular produces [ʃt] coda clusters, should produce fricatives with an increasingly higher center of gravity—becoming more [st] like.

One way to test for phonetic alignment is by exploiting online media archives (e.g., (Gregory and Webster, 1996)). Interestingly, a prominent German politician, Wolfgang Schäuble, PhD, tends to use the Swabian pronunciation of the coda cluster *-st*, while interviewers typically use the standard variant. Making use of large online media corpora, we compared interviews in which a given interviewer interacted with Dr. Schäuble with interviews in which the same interviewer interacted with a speaker using the standard German variant. By comparing the same interviewers over different interviews, we can see whether the interviewee's divergent phonetic choices influence the phonetic choices of the interviewer. Ideally, one would also compare the behavior of the interviewee when confronted with Standard German and Swabian German interviewers. This is, for obvious reasons, not possible, because there are only very few interviewers that use a regional accent. Moreover, it is more likely that the interviewers will align with Dr. Schäuble than vice versa. (Gregory and Webster, 1996) found that social status influences the direction of alignment, so that the interviewer will align with high-status guests while lower-status guests align with the interviewer. Dr. Schäuble is regarded as one of the intellectual heavy weights in German politics and is respected across the political spectrum[1]. It should hence be more likely that the interviewers should align than vice versa. As a test of whether alignment is automatic, we focus on just the way the fricative in a high-frequency word is produced. The logic is that, if alignment is automatic and there is a clear difference in speech gestures (alveolar versus post-alveolar), alignment should occur. It is important to note that a failure to find alignment here does not preclude that other aspects, especially in the domain of prosody, may align. However, in the discussion of the perception-production link, speech gestures have figured prominently (Fowler, 1996; Ohala, 1996), so that a failure to find alignment in this respect is theoretically meaningful.

2 Methods

An online search resulted in seven interviews of more than five minutes with Dr. Schäuble (see Table 1). For each of these interviews, a control interview was identified in which the same interviewer conversed with an interviewee that used the standard variant. All interviewers and control interviewees spoke standard German with no local coloring. Given that the acoustic properties of fricatives are strongly influenced by surround-

[1]To provide one example of this, it is worthwhile to consider the debate about the capital of Germany after the reunification, the so-called Hauptstadtdebatte. In a parliamentary sitting in which there was no voting among party lines, it is widely assumed that Dr. Schäuble's speech tipped the scale in favor of Berlin, even though it had been anticipated that Bonn, the capital of former West Germany would stay the capital after reunification.

ing vowels (e.g., (Smits, 2001)), we focused on productions of the (very) high-frequency word German word *ist* (Engl. 'is'). In each interview, all occurrences of *ist* were annotated if they had a clear fricative portion. Tokens were rejected if the fricative was phonetically voiced or when the fricative was followed by another fricative with no clear boundary[2]. Additionally, up to ten instances of the phonetic string [ɪʃ] and [ɪs] were identified in words that were not associated with any geographical or sociolinguistic differences[3]. These tokens served to indicate whether the fricative in ist is more similar to a speaker's alveolar /s/ or post-alveolar /ʃ/. This comparison is important, because fricative spectra can vary strongly between speakers due to factors such as articulator size and also between recordings because of different levels of ambient and system noise during the recordings as well as the high-frequency cut-off of a recording.

3 Results

First, it was checked whether the control interviewees indeed showed the Standard German pattern with a fricative in *ist* that is similar to their alveolar [s]. This was clearly the case. The control interviewees produced a fricative in *ist* (mean CoG: 5887 Hz) that is similar to their /s/ (mean CoG: 5838 Hz) but differs from their /ʃ/ (mean CoG: 4010 Hz). A linear mixed-effect model with the /s/ mapped on the intercept showed that the fricative in *ist* was not different the other /s/'s ($b_{\text{Fricative="}ist\text{"}} = 46$, t = 0.56), but that the difference between /s/ and /ʃ/ was significant ($b_{\text{Fricative=/ʃ/}} = -1823$, t = 5.57). Having established this, it can now be examined whether the fricative productions from the Swabian speaker Dr. Schäuble indeed deviate from the standard pattern, and whether this in turn influences the interviewers. Table 1 shows the individual means and Figure 1 the overall means for the fricatives' center of gravity (CoG) for tokens of *ist*, tokens of [ɪʃ], and tokens of [ɪs] for the Swabian speaker in different interviews and the respective interviewers' CoGs in conversation with this Swabian speaker or a speaker with the standard pattern. The means show that the interviewers produce a fricative in *ist* that is similar to their [s] in both types of interviews, while the Swabian speakers shows the expected deviant pattern, with a fricative in *ist* that is more similar to the [ʃ].

Table 1: Individual mean spectral centre of gravity for the fricatives by interviewers conversing with a Standard speaker or a Swabian speaker. Note that level difference between different interviews by the same speaker are caused by recording quality.

Interview by (duration in s)	**interviewer/ist/-exposed**			**interviewer/iʃt/-exposed**			**Swabian speaker**		
CK (1786)	5957	5575	3770	6129	5815	3847	4672	6082	4258
DB (1733)	6690	6478	5063	6535	6740	5343	4553	5953	4206
AR (539)	6118	6284	3244	6036	6165	3976	5086	5890	4564
KS (489)	6226	6239	3810	6236	6337	4621	5073	6203	4536
ET (3555)	6184	5716	4136	6370	6078	4004	4378	5600	3954
TJ (2710)	5693	5526	3644	5959	5656	3223	4353	5787	3998
MK (576)	5779	5751	2989	6345	6210	3800	4330	6295	4099
Average	6092	5938	3808	6230	6143	4116	4635	5973	4231

Three linear-mixed effect models were run to test whether there is any alignment of fricative spectra in the data set. A first analysis established statistically that the interviewers and the Swabian speaker differed in their pronunciation of "ist". The dependent variable in this first analysis was the fricatives' CoG predicted by the fixed factors Fricative (three levels: "s", "sch", and "ist") and Role ("Interviewer" vs. "Interviewee") and their interaction. The level "s" for the Fricative factor and the level "Interviewer" for the factor Role were mapped on the intercept. To account for speaker and recording differences, a random intercept was added for each combination of speaker and recording, as well as a random slope for both Fricative and Role. Table 2 shows the resulting beta weights and their level of significance (based on the conservative assumption of 8 df, 14 independent observations minus 6 parameters). Going through Table 2, the Intercept value of 6145 reflects

[2]The German word *ist*/ɪst/ is often pronounced without the final /t/, and the resulting form can be subject to voice assimilation (*ist es*, Engl. 'is it', /ɪst#ɛs/ → [ɪzɛs]) as well as place assimilation (*ist schon*, Engl. 'is already', /ɪst#ʃon/ → [ɪʃ:on]).

[3]Words such as *demokratisch* (Engl., 'democratic') and bis (Engl., 'till') end on [ɪʃ] and [ɪs] respectively in both Standard and Swabian German.

the estimated mean for the combination of factor levels mapped on the intercept, which is the interviewers' average [s] CoG. The beta-weight for the Fricative level [ʃ] indicates that the CoG for [ʃ] by interviewers is more than 2000 Hz lower than their CoG for [s]. The insignificant beta-weight for the "ist" level of the factor Fricative shows that the interviewers produce a fricative in *ist* that is similar to their [s]. Note that the "main effect" for the Swabian interviewee is different from a main effect in an analysis of variance. In a regression model, it shows how the interviewee differs from the interviewers for the level [s] of the Fricative factor, which has been mapped on the intercept. As the estimate shows, there is no significant difference in the pronunciation of [s]. The interviewee produces, however, a slightly higher [ʃ], and, reflecting the Swabian accent, a massively lower fricative in *ist* than the interviewers.

Note that both interviewers and the Swabian speaker seem to produce a fricative in *ist* that is slightly higher than the "reference category" (i.e., /s/ for the interviewers and /ʃ/ for the Swabian speaker). This is likely a residual trace of the /t/ at the end of *ist*, which also has an alveolar place of articulation. Coarticulation with the /t/ would explain the slightly higher CoGs in ist compared to the relative reference category.

Table 2: Beta weights for the analysis comparing fricatives' center of gravity between interviewers and the Swabian speaker.

β	Estimate	t	p (based on df = 8)
Intercept	6145	45.1	< 0.001
Fricative = [ʃ]	-2025	-14.5	< 0.001
Fricative = "ist"	109	1.2	0.14
Role = Interviewee	-173	-1.1	0.16
Fricative = [ʃ] : Role = Interviewee	280	1.6	0.07
Fricative = "ist": Role = Interviewee	-1449	-9.3	< 0.001

Table 3: Beta weights for the analysis comparing interviewers' fricatives' center of gravity when the interviewee uses the Standard or Swabian variant.

β	Estimate	t	p (based on df = 8)
Intercept	6143	44.1	< 0.001
Fricative = [ʃ]	-2026	-13.7	< 0.001
Fricative = "ist"	117	-1.2	0.13
Interviewee = Standard	-210	-1.0	0.17
Fricative = [ʃ] : Interviewee = Standard	-90	-0.3	0.38
Fricative = "ist": Interviewee = Standard	54	0.4	0.34

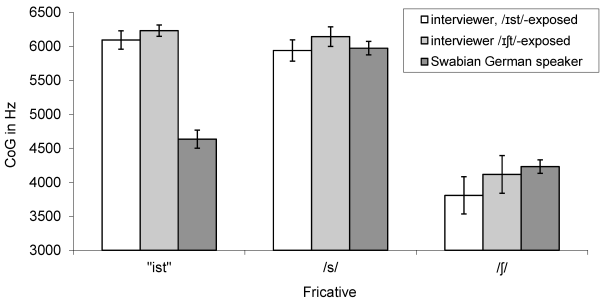

Figure 1: Overall mean spectral center of gravity for the fricatives [s], [ʃ], and the fricative in the German word *ist* (Engl.' is'). Note that centre of gravity of the fricatives in *ist* is similar to [ʃ] for the Swabian German speaker but that this does not influence the interviewers, who produce a fricative in *ist* that is close to their [s] independent of their interviewee's behavior.

Having established an overall difference between interviewers and their Swabian interviewee in the pronunciation of the fricative in *ist*, a next analysis tested whether interviewers produce different fricatives depending on the accent of their interviewee. Again, a linear mixed effect model was run with a random intercept plus random slopes for every combination of speaker and recording. The fixed factors were Fricative and Interviewee's Variant (Standard vs. Swabian) and their interaction. As Table 3 suggests, there was no measurable influence of the interviewees' variant on how the interviewer produces his/her fricatives. First of all, Figure 1 (comparing the white bars for the /ɪst/-exposed with the gray bars for the /ɪʃt/-exposed condition) shows that the observed CoGs in the /ɪst/-exposed condition are all slightly lower than in the /ɪʃt/-exposed condition. This is in all likelihood due to different recording set-ups and audio coding of the AV files for archiving, and is re-

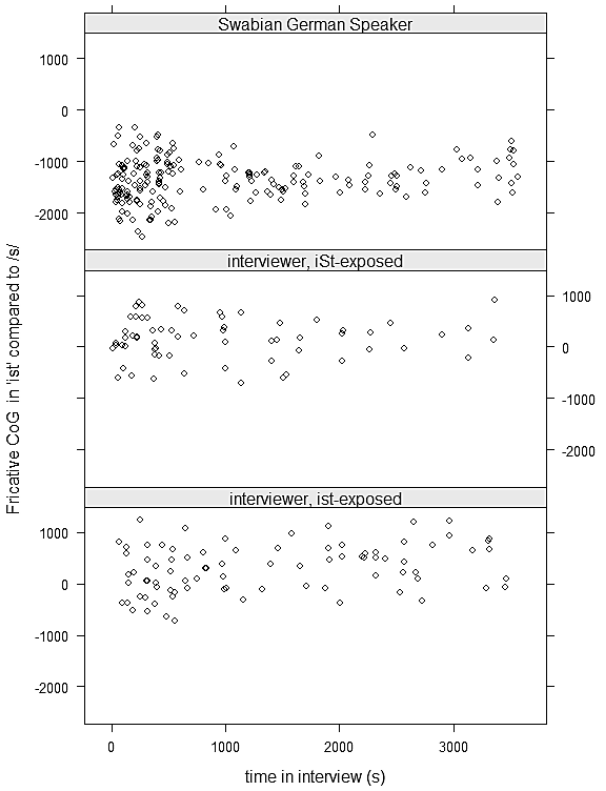

Figure 2: Individual fricatives' centre of gravity in the word *ist* (Engl. 'is', in comparison for the speaker and recording specific [s], plotted against interview duration. Phonetic alignment should lead to an upwards trend for the Swabian speaker and a downwards trend for interviewers when exposed to [ɪʃt] pronunciations. Such effects are neither visible nor found in the statistical analysis.

flected in the -210Hz, but not significant, beta-weight for "Interviewee = "standard". Relative to this baseline difference, the critical interaction term is Fricative = "ist" : Interviewee = "Standard", which indicates how much higher the fricative in ist is if the interviewer converses with a speaker that uses /s/ in *ist*. Note that the effect goes in the "expected" direction, but is, first, far from significant and, second, only about 1/30 of the difference between the speakers; the effect is 54 Hz compared to the nearly 1500Hz difference between interviewer and interviewee in the production of the fricative in *ist*. To confirm that this is no real effect, the initial model with the interaction was compared to the model without an interaction; an analysis that showed that the interaction did not explain any variance ($\chi^2(2) = 0.26$, $p > 0.2$).

Finally, it was tested whether there was any sign of alignment over the course of the interviews. To this end, we generated CoG values for the fricatives in *ist* which were normalized for speaker and recording influences by subtracting the mean [s] value for the recording/speaker combination in which a given token of *ist* occurred. The final regression model then tested whether these normalized values converged over time. The model predicted the normalized CoG of all *ist* tokens with the following fixed factors: Role with three levels ([ɪʃt]-exposed Interviewer, [ɪst]-exposed Interviewer, and Swabian German Interviewee) as well as time in interview. If there is convergence, the fricative CoGs should become higher over time for the interviewee and/or lower for the interviewers when [ʃt]-exposed. The first effect would show that the interviewee aligns with the interviewer while the latter would show that the interviewer aligns with the interviewee. As Figure 2—with all data points for this analysis—suggests, the patterns were stable over time. The beta-weights of the linear mixed effect model shown in Table 4 confirm this. There is no significant interaction of Time and Role. In fact, a regression model with only Role as predictor does not explain less variance than a model with Time and its interaction with Role ($\chi^2(3) = 1.32$, $p > 0.2$). This indicates that the speakers were stable over time.

Table 4: Beta weights for the analysis testing an influence of time on the normalized fricatives' center of gravity in the German word *ist* ('is) for standard speakers under different exposure condition and the divergent Swabian speaker.

β	Estimate	t	p (based on df = 8)
Intercept	207	2.1	0.03
normalized Time	-1	-0.4	0.65
Role = iSt-exposed interviewer	-43	-0.3	0.35
Role = Swabian Interviewee	-1589	-10.7	< 0.001
normalized Time : iSt-exposed Interviewer	-1	-0.2	0.42
normalized Time : Swabian Interviewee	1	0.7	0.24

4 Discussion

The current result indicates that speakers with different accents can maintain their phonetic differences over the course of a conversation. This finding has several theoretical consequences. Phonetic alignment is often portrayed as an automatic consequence of being in a dialogue, with social influences only moderating the inevitable alignment (Miller et al., 2010). The current data set shows that alignment of speech gestures for a

given segment is not inevitable and can be avoided completely.

The contrast of the current study (with no alignment) and other studies (finding alignment at least in perceptual measures), raises the question which parameters are likely to give rise to alignment. First of all, the difference in pronunciation of *ist* in German is well represented in the public conscience, possibly because the difference can be coded orthographically. It might hence be that being conscious of a difference impedes alignment. However, other studies using both acoustic measures and perceptual judgments (Babel et al., 2013; Pardo et al., 2013) also find little alignment of segmental properties such as vowel spectra, which are difficult to capture in orthography, but still find clearer effects in perceptual judgments. These are probably driven by prosodic properties. It may hence be the case that prosodic parameters are more likely to give rise to alignment than segmental properties.

Finding that alignment might not be a consequence of a direct perception-action link suggests that social variables may not be moderators but actually the driving forces of alignment. A similar conclusion is reached by a study (Gregory and Webster, 1996) that analyzed a database of interviews from *Larry King Live*. They evaluated the average spectrum in the band 0-0.5 kHz of different interviewees and Larry King. Based on correlations between different spectra, they argued that interviewer and interviewee accommodate to one another and that who accommodates to who is dependent on the relative social status. Much of their findings, however, are questionable because the analysis was built on wrong assumptions about what drives correlations between spectra[4].

It has also been suggested that alignment fosters mutual understanding (e.g., (Miller et al., 2010; Pickering and Garrod, 2004)). While it is difficult and probably impossible to judge the quality of an interview, listening to the interviews while searching for tokens of *ist* did not suggest that interviewers had trouble in spoken-word recognition caused by the different pronunciation of *-st* clusters by the Swabian interviewee. The current literature on spoken-word recognition indeed suggests that listeners can adapt fast to speaker-specific idiosyncrasies (Clarke and Garrett, 2004; Eisner and McQueen, 2006; Kraljic and Samuel, 2006; Maye et al., 2008; Norris et al., 2003). Quite relevant for the current purposes, (Kraljic et al., 2008) tested adaptation to variation in the pronunciation of fricatives as either /s/ or /ʃ/. They found that quick adaptation in perception did not have any repercussions for production. Similarly, a study by (Mitterer and Ernestus, 2008) suggests that a difference in speech production patterns does not have to hinder perception. In this study, participants had to shadow /r/-initial nonwords in Dutch. The initial /r/ was produced as either an alveolar or a uvular trill, with both variants being common in the Netherlands. Participants did not imitate the variation in the trill; an unsurprising finding as most Dutch speakers master only one trill variant. More interestingly, however, the shadowing latencies were not slowed down by the consequential gestural mismatch between stimulus and response. Participants were just as fast in producing a nonword when the stimulus contained their preferred trill than when it contained the other trill. These datasets seem to converge on the conclusion that a divergence in phonetic patterns does not necessarily impede speech perception and spoken-word recognition. Two interlocutors can agree to disagree on how to produce certain words without negative consequence for mutual understanding.

[4]These authors assume that neither the speaker intrinsic f0 range nor the vocal effort influence the correlation between two spectra. However, a speaker with a average f0 of 130 Hz will have "bumps" in the spectra at the harmonics of the modal f0 which influence the shape of the spectrum. Increasing vocal effort does not only influence the average level of a spectrum, but also changes the spectral tilt and thereby also influence the shape of a spectrum.

References

Babel M., McAuliffe M., Haber G. (2013). Can mergers-in-progress be unmerged in speech accommodation? *Front. Psychol.*, 4, 653.

Brouwer S., Mitterer H., Huettig F. (2010). Shadowing reduced speech and alignment. *J. Acoust. Soc. Am.*, 128(1), EL32–EL37.

Clarke C. M., Garrett M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.*, 116, 3647–3658.

Eisner F., McQueen J. M. (2006). Perceptual learning in speech: Stability over time. *J. Acoust. Soc. Am.*, 119, 1950–1953.

Fowler C. (1996). Listeners do hear sounds, not tongues. *J. Acoust. Soc. Am.*, 99, 1730–1741.

Fowler C., Brown J., Sabadini L., Welhing J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *J. Mem. Lang.*, 49(3), 396–413.

Goldinger S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.*, 105, 251–279.

Gregory S. W. J., Webster S. W. (1996). A nonverbal signal in voices of interview partners effectively predicts ommunication accommodation and social status perception. *J. Pers. Soc. Psychol.*, 70, 1231–1240.

Harrington J., Palethorpe S., Watson C. (2000). Does the Queen speak the Queen's English? *Nature*, 408, 927–928.

Kraljic T., Brennan S., Samuel A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107, 51–81.

Kraljic T., Samuel A. G. (2006). Generalization in perceptual learning for speech. *Psychon. B. Rev.*, 13, 262–268.

Maye J., Aslin R. N., Tanenhaus M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Sci.*, 32, 543–562.

Miller R., Sanchez K., Rosenblum L. (2010). Alignment to visual speech information. *Atten. Percept. Psychophys.*, 72, 1614–1625.

Mitterer H., Ernestus M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168–173.

Nguyen N., Dufour S., Brunelliere A. (2012). Does imitation facilitate word recognition in a non-native regional accent? *Front. Psychol.*, 3.

Norris D., McQueen J. M., Cutler A. (2003). Perceptual learning in speech. *Cognitive Psychol.*, 47, 204–238.

Ohala J. (1996). Speech perception is hearing sounds, not tongues. *J. Acoust. Soc. Am.*, 9, 1718–1725.

Pardo J. (2006). On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.*, 119, 2382–2393.

Pardo J., Jordan K., Mallari R., Scanlon C., Lewandowski E. (2013). Phonetic convergence in shadowed speech: the relation between acoustic and perceptual measures. *J. Mem. Lang.*, 68, 183–195.

Pickering M., Garrod S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.*, 27, 169–226.

Sancier M., Fowler C. (1997). Gestural drift in a bilingual speaker of Brazilian, Portuguese and English. *J. Phonetics*, 25(4), 421–436.

Shockley K., Sabadini L., Fowler C. (2004). Imitation in shadowing words. *Percept. Psychophys.*, 66(3), 422–429.

Smits R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *J. Exp. Psychol. Human*, 27, 1145–1162.