Journal of Phonetics 39 (2011) 18–38

Contents lists available at ScienceDirect

Journal of Phonetics

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

Articulatory events are imitated under rapid shadowing

Douglas N. Honorof[a,*], Jeffrey Weihing[a,b], Carol A. Fowler[a,c]

[a] Haskins Laboratories, 300 George Street, Suite 900, New Haven, CT 06511, USA
[b] Department of Communication Sciences, University of Connecticut, Storrs, CT 06269, USA
[c] Department of Psychology, University of Connecticut, Storrs, CT 06269, USA

ARTICLE INFO

Article history:
Received 14 May 2009
Received in revised form
12 October 2010
Accepted 20 October 2010
Available online 13 December 2010

ABSTRACT

We tested the hypothesis that rapid shadowers imitate the articulatory gestures that structure acoustic speech signals—not just acoustic patterns in the signals themselves—overcoming highly practiced motor routines and phonological conditioning in the process. In a first experiment, acoustic evidence indicated that participants reproduced allophonic differences between American English /l/ types (light and dark) in the absence of the positional variation cues more typically present with lateral allophony. However, imitative effects were small. In a second experiment, varieties of /l/ with exaggerated light/dark differences were presented by ear. Acoustic measures indicated that all participants reproduced differences between /l/ types; larger average imitative effects obtained. Finally, we examined evidence for imitation in articulation. Participants ranged in behavior from one who did not imitate to another who reproduced distinctions among light laterals, dark laterals and /w/, but displayed a slight but inconsistent tendency toward enhancing imitation of lingual gestures through a slight lip protrusion. Overall, results indicated that most rapid shadowers need not substitute familiar allophones as they imitate reorganized gestural constellations even in the absence of explicit instruction to imitate, but that the extent of the imitation is small. Implications for theories of speech perception are discussed.

1. Background

Humans, in some respects exceptionally among primates, imitate one another by reproducing actions and intentions (Galef, 1988; Hauser, 1996; Nagell, Olguin, & Tomasello, 1993; Tomasello, 1996; Whiten & Custance, 1996; but see also Zentall & Akins, 2001). The imitative tendency starts young. Neonates successfully reproduce facial gestures (e.g., Meltzoff & Moore, 1999), and two to five year olds appear to 'overimitate', that is, they reproduce causally or functionally irrelevant aspects of behavior where simple goal emulation would be more efficient (Horner & Whiten, 2005). The imitative inclination continues into adulthood; mature humans appear even to be disposed to imitate emotionally expressive facial gestures of a political leader irrespective of their prior attitude toward him (McHugo, Lanzetta, Sullivan, Masters, & Englis, 1985).

Imitation of speech occurs as well. By twelve weeks of age, infants imitate vocalic sounds (Kuhl & Meltzoff, 1996), and adults are found to imitate speech quite generally. In interactive exchanges, for example, adults are found to converge with other speakers in speaking rate, vocal intensity, accent and other speech characteristics (see Giles, Coupland, and Coupland (1991) for a review of the accommodation literature).

Imitation may serve a variety of functions. For example, Meltzoff and Moore (1999) suggest that imitation provides the foundation on which infants build social cognition. By engaging in reciprocal imitation with people—environmental 'objects' that infants view as most like themselves—infants develop a view of self versus other.

That imitation of speech persists into adulthood, however, invites speculation about other functions. Phonetic convergence between speakers may indicate that imitation marks social affiliation. In fact, discourse topic can, apparently, inspire a small degree of style shifting even when no in-group members are present (e.g., Rickford & McNair-Knox, 1994). Similarly, Bourhis and Giles (1977) found dialect *divergence* on the part of Welsh speakers as they responded to the recorded voice of an Englishman who was disparaging the status of the Welsh language (see also Labov (1963) for a similar finding occurring on a slower time scale).

Although imitation of speech characteristics may mark affiliation in social settings, it occurs in non-interactive settings as well, where its function, if any, is unclear. In particular, it occurs in laboratory settings in which the imitated speaker is present only as a disembodied voice presented by computer as in Goldinger (1998). In that study, Goldinger used a series of immediate and delayed shadowing tasks to test an episodic memory theory that links the incoming signal to stored representations. Goldinger's primary

[*] Corresponding author. Tel.: +1 203 865 8163; fax: +1 203 865 8963.
E-mail address: honorof@yale.edu (D.N. Honorof).

0095-4470/$ - see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.wocn.2010.10.007

measure was perceptual. Listeners heard sequences of three utterances in an AXB paradigm. X was an item produced by a model speaker. A (or B) was the same token shadowed by a different speaker and, therefore, a possible imitation of X. B (or A) was the same item read aloud by the speaker of A (B); thus not an imitation of X. Listeners judged which of A or B was the better imitation of X, and they reliably picked the utterance produced as a potentially imitative repetition of X.

Goldinger's decision to use a perceptual judgment was a wise one for his purposes because there are many dimensions of an utterance, but it may be that only some of them are imitated. Listeners made global judgments as to which of two utterances was more like a model utterance. Yet at times it is also of interest to know which aspects of an utterance are most subject to imitation. For instance, imitation might be of extra-linguistic or prosodic properties only (speaking rate, intonation contour, etc.). Goldinger (1998) did some preliminary acoustic analyses that suggested that duration (or perhaps speaking rate) and fundamental frequency were imitated by his participants. Strong evidence for the influence of fine-grained speech properties on shadowed responses has since been reported (Tilsen, 2009). On a different type of shadowing task, our research group (Fowler, Brown, Sabadini, & Weihing, 2003; Shockley, Sabadini, & Fowler, 2004) found reliable imitation of voice-onset time (VOT). When our participants shadowed words in which the VOTs had been extended, their shadowing responses had longer VOTs than when they shadowed words with unaltered VOTs (Fowler et al., 2003) or than when they named printed words (Shockley et al., 2004). These findings suggest that subphonemic properties of words may be perceived and even imitated in shadowing.

In the present investigation, as in the VOT studies described above, our stimuli involve an adjustment to sounds in the participants' phonological inventory. Here, we investigate shadowing of different types of lateral, which allows for straightforward articulatory decomposition that was not possibly the case in the VOT studies. We ask whether articulatorily manipulated laterals are heard as imitable speech. In Experiment 3, we seek evidence that can tell us whether such laterals, if imitable, are imitated in terms of the gestures manipulated in our stimuli. That is, we ask whether imitations are guided by information extracted by participants about the articulations of the model speaker, as proposed in both the motor theory of speech perception (e.g., Liberman & Mattingly, 1985) and direct realist theory (e.g., Fowler, 1986). In Section 5, we will discuss implications of our research for competing theories of speech perception. We will discuss accounts in which speech is encoded in terms of acoustic (or auditorily filtered acoustic) perceptual objects (see Diehl, Lotto, & Holt, 2004). In addition, we will address accounts which propose that episodic traces of perceived speech utterances underlie perception-driven speech production (e.g., Goldinger, 1996, 1998), and accounts in which perceived speech that affects subsequent production is abstract and phonological (Mitterer & Ernestus, 2008).

The first distinction we examine is related to allophony. Jespersen said that American English laterals, like some British English laterals, are 'dark' in final position and before consonants except /j/. He described these dark laterals impressionistically in terms of a raising of the back part of the tongue toward [u] (§8.6 from a 1969 translation of his original 1912 monograph). In contrast, Jones (1962 revision of 1909) described American English /l/ as being dark in all positions (§302), but did not say whether the 'resonance' associated with American laterals is like [u] as in Received Pronunciation or like [ɔ] or [o], the vocalic 'resonances' he attributed to the dark lateral allophone of London dialectal speech. Indirectly, Delattre (1971) clarified matters by presenting cineradiographic data on prevocalic, postvocalic and geminate laterals across languages. Specifically, he suggested that a pharyngeal gesture for /l/ is always present in American English, but especially noted in non-initial positions. The presence of a salient pharyngeal tongue body gesture even in the initial position might account for why Jones heard American initial /l/s as darker than the British 'light' ('clear') /l/. The relative timing of anterior and posterior gestures for lateral allophones in English is addressed by Sproat and Fujimura (1993) and Browman and Goldstein (1995). The latter in particular report that, in American English, the light variant (henceforth [l]), which occurs in syllable onsets, is produced with a tongue tip closure gesture timed to occur roughly synchronously with a less tightly constricted tongue body gesture. In contrast, the dark variant [henceforth [ɫ]), which generally occurs in syllable codas, is produced with an especially retracted tongue body gesture; the tongue tip gesture, which lags slightly behind the tongue body gesture for [ɫ], may actually undershoot coronal closure. Giles and Moll (1975) discuss a third variant, the syllabic lateral (as in 'apple', 'bottle', 'tunnel', etc.), which they found to have the tongue shape of [ɫ] but the timing pattern of [l]. Gick (2003) compares initial, final and potentially ambisyllabic (in fact, apparently resyllabified) /l/s. In so doing, Gick addresses the notion that, in English, the tongue tip gesture for /l/ counts as a 'C-gesture' but the tongue body gesture for /l/ as a 'V-gesture'. The notion of the two-lingual gestures for a lateral being consonantal or vocalic, respectively, was discussed earlier by Sproat and Fujimura (1993) and Browman and Goldstein (1995), and represents a more precise formulation of the notion of vowel resonances for laterals implicit in work at least as early as that of Jones, noted above. Questions of consonant versus vowel gesture aside, even when one of the pair of gestures is relatively reduced or shifted in time, the sound retains a percept of laterality as attested by over a century of descriptive work on English dialects—a small sampling of which is cited above.

Here, we independently manipulate the two midline constrictions involved in the production of laterals. Specifically, in Experiment 1, we look for acoustic evidence of imitation of /l/ allophones in nonsensical, V.CV sequences when participants are asked to shadow the speech of a model talker rapidly. Because evidence for imitation of this difference is weak in Experiment 1, in Experiments 2 and 3, we modify both types of /l/ by reducing the magnitude of constriction of one or the other of the two midline gestures (tongue tip or tongue body) in an attempt to enhance a perceptible difference between the sounds. Finding stronger acoustic evidence of imitation in Experiment 2, in Experiment 3, we examine articulatory (articulometer) data directly to determine whether acoustic evidence for imitation indicates imitation of underlying gestures consistent with the claims of direct realism and the motor theory.[1]

Acoustic (and auditory) accounts would lead to a different hypothesis from ours. Such theories would lead to the prediction that shadowers who are inclined to imitate will perceive acoustic (or auditory) targets but use whatever articulatory means are necessary to achieve them. Historically, this family of argument arises out of perturbation theory (e.g., Chiba & Kajiyama, 1941). The idea is that multiple articulatory equivalence classes may map onto a single acoustic equivalence class, and that a single acoustic equivalence class (say, [ɫ]) can provide information that allows the listener to recover any one of a number of underlying articulatory configurations. Given this view, from the listener's perspective, all other things being equal, a backing of the tongue body into the area of the vocal-tract-as-tube near the oropharyngeal node (as in [ɫ])

[1] In earlier investigations of speech imitation (Fowler et al., 2003; Shockley et al., 2004), members of our research group found that imitation occurs highly reliably; however, with a small magnitude. The authors have proposed that the small magnitude reflects the fact that most of what guides a listener's production of a word or syllable is his or her own habitual way of producing it. However, that habitual pattern is attracted toward the speech pattern of another speaker.

might be expected to lower F2, while rounding the lips would also lower F2. If so, shadowers might be free to adopt one articulatory strategy or the other with the goal of lowering F2. From our perspective, it seems likely that no matter how many articulatory configurations can allegedly be made to produce a single acoustic pattern in the laboratory, ultimately the mapping between articulation and acoustics must be constrained in the real world to include only the subset of gestural configurations that are anatomically or somatosensorily possible for speakers. The mapping must also reflect the reality that individual talkers show stable preferences for particular possible articulatory configurations (see Bell-Berti, Raphael, Pisoni, & Sawusch, 1979). In theory, it remains possible, however, that individual articulations can be manipulated independently such that only a single formant is affected in a predictable way that does not bear with it a host of other implications for phonatory quality, nasality, airflow, dynamics, etc. For laterals, one such strategy might involve rounding of the lips for [ɫ] to enhance—or even substitute for—the acoustic consequences of tongue retraction. It is important to bear in mind that theories of gesture-based perception do not entail a perfect fit between model utterances and imitated utterances, i.e., our theory does not preclude lip activity for the imitators' [ɫ] in the absence of lip activity for the model's [ɫ]. We predict only that tongue retraction and tip reduction are also imitated in some small way when there is imitation. However, for the sake of thoroughness, we also investigate lip protrusion where doing so is likely to inform one or another theory.

2. Experiment 1

We elicited speech from participants using a variation on a shadowing task implemented by Porter and colleagues (Porter & Castellanos, 1980; Porter & Lubker, 1980) following Koshevnikov and Chistovich (1965). In the task, listeners hear a model speaker producing an extended [ɑ] vowel followed by a consonant–vowel syllable. The participants' task is to shadow the speech they hear by producing the same utterance as the model speaker and by following the speaker as closely in time as possible. Porter and colleagues showed that speakers can shadow model utterances with remarkably short latencies. Our use of the task is different. Our goal is to elicit utterances that are likely to reveal whether differences in articulation are perceived and imitated. By 'differences in articulation' we mean differences of phonetic quality of phone-type that approximate allophonic variation in English, even when positional variation is controlled. Fowler et al. (2003) found that shadowing responses under similar conditions are indeed imitative. We predict imitation here as well.

2.1. Methods

2.1.1. Participants

Eighteen listener-participants were included in the study. By self-report, all were native speakers of American English with no known speech or hearing disorders. Each received $8 per hour for 2 h of participation. The present experiment took about an hour to complete, but participants were run in additional unrelated experiments during the same session.

2.1.2. Stimuli

Model acoustic stimuli were recorded for later presentation to participants as described below. The stimuli were produced by the first author, a native-English speaking phonetician with a complex residence history and complex linguistic background and who was raised in a linguistically complex family. For at least these reasons, his speech patterns may not always be easily identified with any one specific region or social group (see, for example, Payne, 1980), but, in general, the model sounds North American, produces relatively light [l]s in citation form syllable onsets and relatively dark [ɫ]s in citation form codas, and, at the time of the recordings, did not typically vocalize laterals in any position (self-report). All stimuli were vowel–consonant–vowel (VCV) non-words, where the initial and final vowels were always [ɑ] and the consonant was [l], [ɫ], /r/, or /w/.

The model was asked to produce VCVs whose initial vowel varied between 2000 and 5500 ms in increments of 500 ms. This yielded a total of eight different initial vowel durations: 2000, 2500, 3000, 3500, 4000, 4500, 5000 and 5500 ms, approximately. The model was prompted via a computer terminal so that he would know when to begin and end the production of the initial vowel. Varying the initial vowel duration served to prevent participants from predicting the moment of consonant closure and was not a factor in later analyses.

At least three recordings were made of every possible duration-by-consonant combination ([l], [ɫ], /r/ and /w/) for a total of 24 tokens per consonant. Here /r/ and /w/ tokens were originally included in the experiment only to make it more difficult for the participants to guess the nature of the experimental manipulation, though analysis of /w/ did become relevant to theoretical concerns addressed in the end.

Model acoustic stimuli were recorded digitally at 20 kHz using a shotgun microphone with a 50 Hz–20 kHz ± 3 dB frequency response. All recordings were filtered with the Spark XL denoising algorithm (TC Electronics Inc., Westlake Village, CA) to remove any electronic or ambient noise in the signals that may have proven distracting to shadowers. A comparison between filtered and unfiltered stimuli revealed no noticeable adverse effects in the region of the relevant formants.

Physiological data were acquired simultaneously using a midsagittal magnetometric system (Perkell et al., 1992). The model's articulations were collected to provide a basis for excluding tokens in which the model failed to produce /l/ variants according to the experiment instructions. No articulatory data were collected from the participants, nor did participants speak with coils affixed to their articulators. Tokens were excluded from the stimulus set when no measurable gesture was found in the vertical movement of the model's tongue tip. Similarly, tokens in which there was no identifiable horizontal movement of the tongue body were also excluded. Altogether, we excluded three tokens: one [ɫ] with an initial vowel duration of 5000 ms and two [ɫ]s with an initial vowel duration of 5500 ms. Valid tokens with identical duration-by-consonant combinations were repeated a sufficient number of times to compensate for the excluded tokens.

In investigating the timing of achievement of target for lateral tongue tip and tongue dorsum gestures, Gick (2003) reports a trend in the direction of negative tip-lag for syllable-initial laterals but positive tip-lag for both syllable-final and syllable-final laterals that are potentially resyllabified to syllable-initial or ambisyllabic position (such as laterals that occur before a vowel in connected speech). We note that the /l/-producer whose articulation was investigated by Gick spent his formative years in the same geographic region as did the model. In the present study, however, informal evaluation of the model's remaining articulatory data revealed that achievement of target for the tongue tip closure gesture was timed roughly synchronously with the achievement of the tongue body retraction gesture in both conditions, not just in the [l] condition, a finding that suggests that the model was not simply substituting his typical allophone in either case. Nevertheless, our informal observations of the model's laterals indicate approximately 60 ms (on average) greater lag time for [ɫ] than for [l] between the onset of retraction of the tongue body coil and onset of raising of the tongue tip coil in the direction of the palate as indexed by sudden increases in articulator velocity out of a period of near-stationarity. This onset-to-onset lag pattern suggests that the two /l/ sounds were not timed identically and that the tip-lag seen in the model's articulatory onsets follows the direction of tip-lag reported elsewhere in achievement of targets for [l] and [ɫ] in more typical syllable positions. The lightness of the model's [l]s and the darkness of the model's [ɫ]s are confirmed in the acoustic measures reported below. Therefore, we conclude that the model was successful in following instructions to produce the gestures associated with [l]s and [ɫ]s, respectively. Given a near-universal preference for dividing ambiguously syllabified VCV sequences into V.CV or perhaps producing truly ambisyllabic consonants, the timing of the model's [ɫ] gestures appears to have been adjusted (appropriately) to accommodate the intervocalic context; the [ɫ] retains positive tip-lag in its onsets even in a context that allows for resyllabification. Furthermore, any listener bias introduced by the abnormally long durations of the preceding vowel would be expected to lead shadowers toward, not away from, V.CV syllabification. The lateral variants were otherwise unremarkable.

As noted above, in the present experiment, by design, all /l/s were embedded in a non-word V.CV context for purposes of comparison. Where [l] was intended, if the shadowing task predisposed subjects to impose syllable-affiliation judgments, the [l] would naturally be heard as an unremarkable syllable onset [l] in that such relatively light [l]s are typical in American English onsets. Owing to the modest magnitude of the onset-to-onset tip-lag in the [ɫ]s in the model's productions, participants may have heard the [ɫ] as unremarkable (that is, a resyllabified [l]). If the magnitude of the tongue body retraction or the extent of the onset-to-onset tip-lag was sufficient to trigger the percept of an [ɫ], the mismatch of chromatics and position might have predisposed our listener-shadowers to perceive the [ɫ] in the present experiment as a juncture geminate (that is, what we expect in connected speech where one word ends in an /l/ and the next word begins with one) or perhaps like the velarized onset [ɫ] heard in some familiar accents of English (Lunn, Wrench, & Mackenzie Beck, 1998; Wells, 1982: 411–12l).

2.1.3. Procedure

Data were collected from participants in two blocks. A block consisted of a randomized presentation of 24 [l], 24 [ɫ], 24 /r/ and 24 /w/ tokens. Each of the 24 /r/ and /w/ tokens was presented twice per block so that participants would not detect a greater abundance of /l/s.

Written instructions for the speeded shadowing task were given to participants to highlight the most important aspects of the experiment (see Appendix A). Participants were told they would hear a 2000 ms (sine wave) warning tone over headphones after which they would hear the model produce a sustained vowel leading directly into a consonant–vowel syllable. They were asked to repeat what they heard the model saying, and to keep up with him as closely as they could. Trials were self-paced. Participants were instructed verbally to inform the experimenters of any speech errors before moving on to the next stimulus.

Generally, participants were not informed of the specific consonants they would be hearing. The exception, however, was the consonant /r/, to which they were exposed twice prior to beginning the experiment. The first exposure to /r/ occurred in the written instructions, where an /r/ example was presented. The second occurred when participants were presented with three acoustic examples of good shadowing. The examples consisted of playing a recording of the model producing a VCV through one speaker of the headphones, followed by a recording of one experimenter shadowing the model closely through the other speaker. Given that no analyses relating to /r/ were conducted, advance exposure to this particular consonant is assumed to be inconsequential.

Participant recordings were made in a sound isolation booth with an omni-directional microphone having a very flat 4 Hz–40 kHz ± 1 dB frequency response. All audio recordings were initially sampled at 44.1 kHz. However, prior to analysis, they were down-sampled to 22,050 Hz using SoundApp 2.6.1 (Norman Franke, Livermore, California, USA). An 80 Hz hardware high pass filter (M80, PreSonus Audio Electronics, Baton Rouge, Louisiana, USA) was applied at the time of recording.

2.1.4. Acoustic label placement

Shadower formants were tracked by Linear Predictive Coding in Macquirer 4.9.7 (Scicon R&D, Inc.) with 26 LPC coefficients (frame length=256 samples; step size=221; smoothing: 5–7 sample rectangular window). A label was placed at the F1 minimum corresponding roughly to the temporal midpoint of the /l/, and the F1 and F2 values at that frame were logged.[2] If there were multiple samples sharing the minimum F1 value, the label was placed at the F1 minimum that had the lowest corresponding F2 value.

For the model, despite multiple adjustments to frame size and other LPC parameters, the analysis failed to produce consistently interpretable results for laterals due to apparent antiresonances. Therefore, to reduce effects of window position and of the varying characteristics of the vocal tract during the open and closed phases of glottal vibration, we shifted to the more accurate technique for extracting the resonances of the vocal tract given in Yegnanarayana and Veldhuis (1998). First we estimated the quasi-periodic instants of significant excitation corresponding roughly to the instants of glottal closure for voiced speech segments. These instants were identified at the positive zero crossings in the phase-slope function (computed from short-time [20 ms] spectrum analysis and the average group delay). Next we chose very short analysis windows (less than a pitch period) and synchronized them around the instants of significant excitation that we identified automatically. Ideally such regions included the more stable, less damped, post-excitation phase of each glottal cycle. The post-excitation phase is believed to correspond to the closed phase of the glottis during phonation of voiced sounds such as /l/. This phase was chosen to reduce the effect of coupling of the free resonances of the vocal tract with the contributions of the trachea, air flow and air pressure that occur in the open phase region of the glottal cycle. We obtained the complex poles by computing the roots of the prediction polynomials derived from these short window regions, applying a

[2] Because LPC analysis assumes that "the voicing spectrum is primarily shaped by broad spectral peaks with no prominent spectral valleys" (Johnson, 1997b:87), formant tracking is not expected to completely represent the spectra of laterals. Indeed, in the present study, there were cases where the presence of antiresonances as seen in a broadband spectrogram made it difficult for the LPC analysis to trace F1 during the /l/ transition. In such cases, the LPC trace contained discontinuous values. Traces that were comprised primarily of discontinuous points during the /l/ were not measured. However, there were tokens in which discontinuous values occurred in only a small region of the formant trace during the /l/. It was possible to locate a valid F1 minimum in such cases by following one of two procedures. The specific procedure for labeling these cases depended on which pattern of discontinuous values was found.

In the first pattern, discontinuous values were isolated and grossly disconnected from the F1 trace. In other words, there was a large jump in frequency between the curve frames during the lateral and a spurious F1 value, with no values occurring in between. The solution in these cases was to place the label at the F1 minimum, excluding any spurious value from consideration.

In the second pattern, there were obvious spurious values, but they were not as isolated as in the first pattern; there was a break in the steady state trace with a large step down in frequency, followed by a gradual rise back toward the valid steady state as judged by eye with reference to the formant pattern seen in a broadband spectrogram. Here we placed the label on the sample having the lowest valid F1 minimum within the steady state.

multi-cycle covariance method to compensate for the shortness of the analysis windows and for the effect of turbulent and other noise in the glottal waveform itself. No formant smoothing was applied. A frame size of 5 ms was used throughout the analysis. Finally, we obtained the temporal frame index corresponding to the lowest value of the F1 formant track during the lateral in each signal. F1 and F2 values at this temporal index were logged by algorithm. Spot-checking of these measurements indicates that they are accurate to within one analysis frame.

All acoustic analyses of model and participant productions were conducted on the transformed variable, F2–F1. Sproat and Fujimura (1993) have shown that F2–F1 distances are smaller in [ɫ] than in [l]. It may be that the smaller F2–F1 value for [ɫ]s is primarily a function of light/dark differences in F2. Narayanan, Alwan, & Haker (1997) speculate that F2 "can be associated with the half-wavelength resonance of the back cavity…Retracting or raising the posterior tongue body observed in the case of [ɫ] results in an increase in the effective length of the back cavity, and hence a lowering of the F_2 values" (p. 1074).[3] Therefore we may expect values of F2–F1 to be smaller for [ɫ] than for [l], just as they were in Narayanan et al.'s study.

2.2. Results

2.2.1. Acoustic data exclusion

Cases were classified as speech errors if the participants did one of the following: indicated verbally after shadowing that he or she made a mistake, ceased shadowing the initial vowel prior to the CV syllable and did not resume shadowing, hesitated between the production of the initial vowel and the CV syllable, clearly produced a non-speech sound (e.g., a cough), and/or uttered one CV syllable, but then abruptly switched to a different CV syllable. No CV syllables were excluded on the basis of judgments of quality of the match with the model target, nor did we exclude utterances in which multiple consonants were produced during closure (e.g., CCV) unless they were specified as errors by the participant. Using these criteria, we determined that 1.9% of the participants' [l] data and 4.5% of [ɫ] data were speech errors.

There were three additional criteria for removing data. First, there were some responses that, although they were not classified as speech errors, could not be measured because of difficulties encountered in applying LPC formant tracking. 6.4% of the [l] and 8.2% of the [ɫ] data were removed due to this type of estimation uncertainty. Second, a very small percentage of the responses were removed because of acquisition error such as a truncated waveform, resulting in the exclusion of fewer than 1% of the [l] and [ɫ] data. Third, data that were 2.5 standard deviations above or below the mean of the group were deemed outliers and were excluded from the data set. Outlier removal accounted for 3.0% of the [l] and 3.7% of the [ɫ] data. Overall, approximately 12% of the [l]s and approximately 17% of the [ɫ]s were removed for one or another of the aforementioned reasons.

2.2.2. Acoustic analysis—ANOVAs

The model's F2–F1 data were entered into an ANOVA with the independent variable '/l/ type' ([l] vs. [ɫ]). There were 24 [l] and 21 [ɫ] tokens included in the analysis. The main effect was significant ($F(1,43)=36.37$; $p<.001$), with a greater F2–F1 mean value for the [l] ($M=584.58$; SD=117.5) than for the [ɫ] ($M=385.32$; SD=102.03), a difference of approximately 199 Hz. This test confirmed that the model was producing /l/ variants as instructed.

Participants' F2–F1 averages were entered with the same independent variable into a repeated measures ANOVA ($N=18$). The main effect was significant ($F(1,17)=6.31$; $p<.03$), with a greater F2–F1 mean value for the [l] ($M=806.83$; SD=161.42) than for the [ɫ] ($M=786.32$; SD=169.10), but a difference of only approximately 20 Hz. The direction of the participants' [l]–[ɫ] mean difference was consistent with the direction of the model's mean difference.

2.3. Discussion

The first experiment provided statistically reliable evidence for imitation as we had expected based on earlier findings (e.g., Fowler et al., 2003; Goldinger, 1998). Like the model, participants showed a significantly smaller F2–F1 difference for [ɫ] than for [l]. Participants did not merely substitute a single type of /l/ for both variations. Light/dark distinctions between /l/-types were imitated, and therefore, we conclude, must have been perceived. However, participants did not produce nearly as great a difference between types of /l/ as did the model. Whereas the model showed an F2–F1 difference of approximately 199 Hz, participants showed a modest 20 Hz difference. No special explanation for the smallness of the effect magnitude is needed; as we noted earlier, the magnitude of imitation tends to be small in rapid shadowing when no instruction to imitate is given (see, for example, Fowler et al., 2003).

However, it is necessary to address the question of how the nature of the stimuli and the nature of the task may have led to imitation of differences of /l/ type in a way that might not reflect the normal pattern of speech perception outside the laboratory. Our shadowers heard non-words, which may have caused our listeners to focus unnaturally on the phonological properties of the utterance. In our case this was by design; we would not have known whether an unusual /l/ type was perceived if talkers simply accessed and reproduced their own versions of real words.

More relevantly, one of our stimulus types, while typical phonetically for our talkers, was atypical phonologically in the sense that the 'dark' lateral was presented in an atypical syllable position. That is, the model was instructed to produce both [l] and [ɫ] in syllable-initial position—a position in which [ɫ] does not normally occur in most US accents (but see Gick (2003) for a discussion of resyllabified English laterals). Because there is no listener-independent test of syllable affiliation available to us, we cannot reliably quantify the model's success in this endeavor, but assert strongly that to the ears of all three of the authors (each a native speaker of US English), all stimuli in the present experiment sounded unambiguously to be syllabified as intended, that is, V.CV. It seems to us highly unlikely that our shadowers perceived the syllable boundary differently than we do. Their perception of chromatics is not likely to arise out of syllable affiliation; that is, listeners did not simply perceive V.CV when hearing [l] but VC.V when hearing [ɫ], and then infer chromatics accordingly. Certainly it has been reported elsewhere that a proper match between allophone and allophonic context or conditioning environment facilitates discrimination, perception, lexical access and fidelity of imitation (at least for real words—see Gaskell & Marslen-Wilson, 1996; Whalen, Best, & Irwin,

[3] To avoid confusion, we point out that only two of Narayanan et al.'s four subjects showed so-called velarization (that is, a raising of the tongue dorsum) for [ɫ], which is why they say "retracting *or* raising" [emphasis our own]. All four of their subjects showed tongue root retraction, though only two showed tongue dorsum retraction. One subject in Giles and Moll's (1975) study also velarized, but the other two actually lowered the tongue dorsum. All three retracted the tongue dorsum (p. 213). Given the stability of tongue body backing for darker laterals across studies, in our Experiment 1, the model understood that he was to retract the back of the tongue into the oropharynx for the [ɫ] without concern for which part of the superior surface of the tongue made the tightest constriction. Certainly our articulatory measures were of a flesh point anterior to the actual constriction location. In Experiment 3, at least, in the process of backing the key part of the tongue body, the more distal anterior flesh point was also lowered away from the palate by the model (see Fig. 2).

1997). Our match was improper, but imitation nevertheless obtained.[4]

On the other hand, the fact that the model produced [l]s and, critically, [ɫ]s in syllable onsets leads to a different obstacle to straightforward interpretation of our results. As one reviewer pointed out, it may be that our listeners became aware of the contrast between /l/ types because the darker phone stood out as being positionally abnormal. Combined with the impoverished context of the stimuli (only four types of consonant in one vowel context in non-words), the strangeness of the task may have led our shadowers to attend to the details with enhanced sensitivity. It is known that experimental designs that draw participants' attention to or away from phonology can affect outcomes in perception tasks (for data and a review, see Cutler, Mehler, Norris, & Segui, 1987). Here, the idea would be that because one-sixth of the syllables presented (24 out of 144) contained a sound ([ɫ]) presented in a syllable position where it is not normally found in the model's and shadower's accents (syllable onset), our shadowers were primed to imitate gestures that they otherwise would not have perceived. In other words, it is possible that the relative familiarity of shadowers with the [l] in a V.CV context compared with shadowers' relative unfamiliarity with the [ɫ] in that same context may have caused them to perceive and reproduce gestures in a way that does not inform us about normal speech perception.

We acknowledge that our task is different from the task of perceiving real words in context outside the laboratory. Our experiment shares this trait with most speech perception experiments, which indeed draw attention to phonology in some way. However, we find it very unlikely that this difference would lead perceivers to extract information about phonetic gestures that they might not normally perceive. In any case, the stimuli of Experiment 1 would not likely seem strange to our particular population of shadowers who were all surely previously exposed to accents of English that have syllable-initial dark laterals. To expand on this point, throughout the USA, it is not entirely uncommon for English-speaking children to vocalize laterals (that is, to pronounce /l/ as a vowel), lateral vocalization often being regarded as a cause for clinical intervention. Furthermore, the use of reduced-tip laterals in coda position is a common feature of many accents of non-disordered English (some African-American and Estuary English, notably). Indeed, lateral vocalization can occur even in initial position in some varieties of English. For instance, a vocalized /l/ occurs without regard for syllable position in parts of Scotland and for some English speakers in Australia and New Zealand. Wells describes the vocalized lateral sound in all three regions as pharyngealized or as possibly pharyngealized (1982: 411, 603, 609). It is very unlikely that our listeners have never been exposed to these accents. Even more clearly relevant to our experiment, Faber (1989) points out that tip-reduced laterals occur commonly even in syllable onsets in the speech of many English-speaking natives of New York City. All three of the present experiments were run in New Haven, Connecticut, which lies at the northeast terminus of Manhattan commuter rail service and within the program delivery area for some New York City radio and television stations. It may even be the case that some of the participants in Experiment 1 produce dark laterals in syllable onsets themselves. (We actually prescreened for this accent feature in Experiment 3 with the aim of excluding participants whose laterals sounded vocalized to the experimenters, or whose syllable onset laterals sounded dark. None presented, however.) Whether or not the imitators in Experiment 1 were themselves producers of dark laterals in V.CV position (resyllabified or otherwise), they would have been exposed to such a speech pattern through the international dissemination of recorded media prevalent in recent decades, if not through contact with the many transplanted New Yorkers in the area. Thus, the positional manipulation of the lateral allophones of Experiment 1 would not have struck the participants as universally unusual, and certainly not as non-speech. An unremarkable reaction to the stimuli would be especially likely given any priming effect induced by our experimental instructions which suggested that participants would be hearing speech (see Appendix A).

In Experiment 1, our participants were asked to rapidly shadow nonsense words that had familiar sounds in a contrived context. It would not be fair to conclude that our shadowers were imitating gestures, however. Rather, they were presented with bigestural constellations that correspond to the familiar allophones of their own systems, and are likely to have been imitating by producing bigestural constellations in which the difference between light and dark was somewhat attenuated by the introduction of the oddly syllabified target [ɫ]. In Experiments 2 and 3, we pushed the boundaries further, asking whether rapid shadowers can imitate gesturally simplified laterals that do not correspond to the allophones of their own systems. Specifically, in Experiment 2 we attempted to enhance the salience of the difference between participants' imitations of /l/ variants by increasing the difference in the model's /l/ variants. To this end, the model reduced one or the other midline constriction for the lateral, and participants were asked to shadow him. Acoustic measures were made. In Experiment 3, with the same intended stimulus design, we examined model and shadower articulations more directly via a magnetometer. In the latter two experiments, given that the model fully articulated only one midline gesture per lateral, if listeners are able to perceive the individual gestures, there is no reason to expect them to adopt a strategy of substituting a familiar bigestural constellation based on closest acoustic match; a finding of substitution would suggest that listeners perceive in terms of abstract linguistic patterns (however encoded), which would weaken our theory that listeners perceive gestures directly. We also consider evidence for or against a strategy for dark /l/ that targets F2-lowering rather than tip reduction, specifically one in which the shadower recruits the lips.

3. Experiment 2

In the second experiment, we increased the difference between /l/ types in the stimuli, asking the model to produce 'lighter' [l]s and 'darker' [ɫ]s via reduction of gestural magnitude.

3.1. Methods

3.1.1. Participants

Fourteen participants took part in the experiment. Two additional participants were tested but data collected from them were excluded from the acoustic analysis because fewer than 50% of their [l] or [ɫ] tokens could be analyzed. Participants were self-reported native speakers of American English with no known speech or hearing disorders. They received $8 an hour for 2 h of participation. Like Experiment 1, the present experiment took only 1 h to complete, but participants were run in additional unrelated experiments during the same session.

[4] An anonymous reviewer suggested that neural adaptation to the preceding /a/ context may have enhanced shadowers' sensitivity to lateral acoustics (see Holt, Lotto, & Kluender, 2000). In other words, the formants of /a/ are very different from the formants for either type of lateral presented, so the laterals may have been scrutinized especially carefully by the shadower. We point out that, as one moves from a vowel into a consonant, there are often gross changes to which spectral bands constrain the greatest acoustic energy, so there is nothing remarkable about our stimuli in this respect—nothing that would affect whatever claims the data may support. We certainly acknowledge, however, that laterals would be expected to sound more different from /a/ than from /ow/, for instance.

3.1.2. Stimuli

Stimuli for Experiment 2 were created roughly as for Experiment 1 except that here no physiological data were collected from the model speaker. Acoustic stimuli were recorded directly to hard disk at 44.1 kHz in a sound isolation booth using an omnidirectional microphone. The frequency response of the microphone was 4 Hz–40 kHz ± 1 dB. However, prior to analysis, the recordings were downsampled to 22,050 Hz using SoundApp 2.6.1. An 80 Hz hardware high pass filter (M80, PreSonus Audio Electronics, Baton Rouge, Louisiana, USA) was applied at the time of recording.

The phonetician-model of Experiment 1 again recorded VCV nonwords. The consonants recorded were [l], [ɫ], /r/ and /w/. Although the phonetician was asked to produce /r/ and /w/ in a manner typical of his native accent as in Experiment 1, his goal here was to produce /l/s in a manner *not* typical of his native accent. Specifically, the model's goal was to de-emphasize the retraction of the tongue body for [l] tokens to make them sound 'lighter' than the [l]s from Experiment 1. For the [ɫ] variant, the model's goal was to de-emphasize the tongue-tip gesture while nonetheless retracting the post-dorsal region of the tongue midline into the oropharynx, without making medial contact with the rear wall of the pharynx (see Gick, Kang, & Whalen, 2002). In both cases, the model was to produce a sound easily recognizable as a lateral. Because the model was an experimenter and therefore aware of the motivation behind exaggerating the difference between /l/ types, it is possible that he unwittingly did more than instructed to exaggerate that difference. For instance, he may have shifted constriction location in an unknown way. Unfortunately, we cannot know exactly where the tightest point of constriction occurred because we have no articulatory data on the stimuli for Experiment 2. Nonetheless, we do have articulatory data for the same model's production of stimuli under the same instructions for Experiment 3. Even in Experiment 3, we do not know the location of the posterior soft-palate or pharynx wall in our coordinate space, and have near-dorsal data only on the articulator coil that indexes (in our terms) the Tongue Body (TB), but that coil is not, in fact, affixed to a particularly posterior dorsal flesh point. Therefore, we cannot surmise much about exact tongue body constriction location, only that the model succeeded in de-emphasizing the tip/body gesture for /l/. While the model's goal in Experiments 2 and 3 was not to reduce one or the other gesture to zero magnitude, magnetometry and listener-debriefing in Experiment 3 suggest that he may have done so for both varieties of /l/ at least in that experiment. We safely assume he was at least capable of having done so here as well. Given the model's success in reducing the magnitude of the tongue tip gesture for [ɫ], it is not surprising that, to the model's own ear, the reduced-tip stimuli sounded vocalized to him in both Experiments 2 and 3.

Henceforth we refer to the /l/ variants in Experiments 2 and 3 simply as [l] and [ɫ], though the reader should bear in mind that these transcriptions are very broad in the sense that these same symbols were used in describing a less extremely articulated distinction between '/l/ type' in Experiment 1.

3.1.3. Procedure

The instructions (see Appendix A) and procedures were identical to those of Experiment 1. The /r/ tokens presented as examples of good shadowing were identical as well. In locating events for /l/s, labels were placed on the basis of Macquirer formant-tracking as in Experiment 1. (26 LPC coefficients; frame size = 256 samples; overlap = 221 samples; smoothing = 5–7 sample rectangular window.)

3.2. Results

3.2.1. Acoustic data exclusion

Acoustic data were analyzed for both model and participants using LPC analysis with parameters described for participant data analysis under Experiment 1. Criteria for classifying errors and outliers were also the same as those followed for Experiment 1. Using these criteria, 4.3% of the [l] and 5.4% of the [ɫ] data were removed due to participant error; 7.3% of the [l] and 6.7% of the [ɫ] data were removed due to difficulties encountered in applying LPC formant tracking; fewer than 1% of the [l] data and fewer than 1% of the [ɫ] data were removed due to data acquisition error; and outliers accounted for 2.9% of the [l] and 2.2% the [ɫ] data. Overall, approximately 15% of the [l]s and approximately 15% of the [ɫ]s were removed for one or another of the aforementioned reasons.

3.2.2. Acoustic analysis—ANOVAs

An ANOVA was conducted on the model's F2–F1 data with /l/ type ([l] vs. [ɫ]) as the independent variable. There were 24 [l] and 24 [ɫ] tokens included in the analysis. The main effect was significant ($F(1,\ 46)=347.41$; $p<.0001$), with a greater F2–F1 value for the [l] ($M=590.79$; SD $=47.91$) than for the [ɫ] ($M=332.04$; SD $=48.27$), a difference of approximately 259 Hz, by design, a greater difference for the model than found in Experiment 1.

Participants' F2–F1 averages were entered with the same factor into a repeated measures ANOVA ($N=14$). The main effect was significant ($F(1,13)=13.36$; $p<.003$), with a greater F2–F1 mean value for the [l] ($M=666.97$, SD $=93.47$) than for the [ɫ] ($M=601.13$, SD $=104.05$), a difference of approximately 66 Hz—about three times the difference found in Experiment 1. The direction of the participants' [l]–[ɫ] mean difference was again consistent with the direction of the model's mean difference.

3.3. Discussion

We were successful in replicating our finding of Experiment 1; participants showed a significant tendency to imitate the model even though the model's darker [ɫ]s were in a potentially ambisyllabic or re-syllabifying syllable position where participants would normally produce relatively light laterals of one kind or another. The magnitude of the acoustic difference between /l/ variants for the participants was larger here than it was in Experiment 1 (66 Hz here versus 21 Hz in Experiment 1).

Although the average difference in formant distance by lateral type was several times larger here than it was in Experiment 1, in neither experiment was it close to the difference exhibited by the model; the direction of difference followed that of the model, but not the magnitude of difference. We have seen this before, using three quite different experimental tests of imitation of VOT (Fowler et al., 2003; Sancier & Fowler, 1997; Shockley et al., 2004). That is, in the previous studies, participants' VOTs approached the VOTs of a model speaker or speakers in direction, but were not encompassed within the model's VOT range. We ascribe this pattern to two competing tendencies. One is the disposition to imitate (even without being instructed to do so explicitly) on which the present experiments and the earlier experiments focus; the second is the tendency to persist in habitual ways of producing phonetic segments. This is also consistent with the pattern of results reported for subphonemic vowel priming in Tilsen (2009).

4. Experiment 3

Participants' shadowed productions of the [ɫ] variants in Experiments 1 and 2 were measurably darker than their [l]s (that is, showed closer formant distances) as was the case for the model. While participants did not simply substitute [l] for both variants on the basis of positional information, in both experiments, the imitation evidenced was small in magnitude compared to the acoustic distinction between /l/ variants made by the model.

In order to learn whether the participants really shadowed the model's articulation, or whether they achieved the apparent but minimal imitation effects on some other articulatory basis, we ran a third experiment in which we investigated articulation directly.

4.1. Methods

4.1.1. Participants

Five participants took part in Experiment 3. Data from one additional participant were excluded because fewer than 50% of her [l]s could be analyzed acoustically. All participants were self-reported native speakers of American English of normal speech and hearing. Participants were paid at the rate of $20 an hour for their participation. The average session lasted about 3 h, including time spent on pre-screening, data collection and debriefing.

To make sure it was possible for shadowers to substitute the most similar allophones from their own inventories if they were inclined to do so, steps were taken to ensure that speakers had two allophones to begin with. In preparation for Experiment 3, we ran informal pre-screening on all participants. Specifically, each prospective shadower was asked to read aloud a sentence crafted to require production of laterals in initial, medial, final and potentially resyllabified loci. The sentence, also peppered with /r/ and /w/ distractors, was, "Len Peters says we'll have to let the happy children ride the little red wagon and all the light brown ponies down the great big hill to the lake in the meadow." The first author listened while each prospective shadower produced the sentence five times. By this procedure, we verified that each prospective shadower produced noticeable light–dark variation in the expected syllable positions outside of the experimental context and that the participant did not vocalize coda [ɫ]s noticeably. All prospective shadowers passed the screening.

4.1.2. Stimuli

The method used to generate stimuli for Experiment 3 differed from that of Experiment 2 in that, in Experiment 3, we acquired magnetometric data. The magnetometer was employed so that model and participant acoustics could be compared with meaningful reference to articulation. The model's acoustics were recorded simultaneously during the magnetometer session and were used as stimuli for the present experiment.

For both the model and the participants, transducer coils were placed at six locations on the face and tongue, including: reference coils on the bridge of the nose and the border of the maxillary incisors and gums, coils on the vermilion border of the upper and lower lip, and coils as close as possible to the tongue tip and as posterior as possible on the tongue body—somewhere between the tongue center and tongue dorsum. The model's tongue tip coil to tongue body coil distance was approximately 4.2 cm. Each coil was placed so that its longer dimension was perpendicular to the midsagittal plane.

After affixing the nose and maxilla coils, but prior to attaching any articulator coils, occlusal bite angle data were obtained using a bite plate. Two coils were attached to the bite plate, one inside and one outside the area of dental contact. This information allowed physiological data to be rotated and thereby brought into conformity with the occlusal plane prior to analysis as in Westbury (1994).

Once all coils had been affixed, a palate trace was acquired as the participant slid the tongue tip coil along the midline of the hard palate. The midsagittal curve of the palatal arch was determined on this basis.

The initial and final vowels of the VCVs were always [ɑ], and the consonants were either [l] (tongue body gesture reduced), [ɫ] (tongue tip gesture reduced), /r/ or /w/. The phonetician-model was asked to produce all consonants exactly as he produced them in Experiment 2.

The design for recording stimuli in the present experiment was identical to the design for Experiment 1. Three recordings were made of every duration-by-consonant combination, for a total of 24 tokens per consonant. We decided after these recordings were made, however, that asking participants to shadow the initial vowel for as long as 4000–5500 ms might introduce a problem. Namely, having to shadow a long initial vowel might cause a participant to run out of breath before beginning to shadow the CV syllable. In fact, in Experiments 1 and 2, 10% of the errors were loosely attributable to participants approaching functional residual capacity (that is, end-tidal volume),[5] and the majority of these errors (62%) were on trials with initial vowels greater than 3500 ms. Therefore, only stimuli with initial vowels less than or equal to 3500 ms were presented for shadowing.

Stimuli were recorded directly to hard disk at a 16 kHz sampling rate using a full-condenser shotgun microphone (frequency response: 50 Hz–20 kHz ± 3 dB). Stimuli were denoised with the Spark XL denoising algorithm to remove background noise. A comparison between filtered and unfiltered stimuli revealed no noticeable adverse effects in the region of the relevant formants.

All articulatory data were acquired at a 500 Hz sampling rate. Movement curves were low-pass filtered with a 9th order Butterworth filter twice: once prior to spatial recovery of the voltage data, and once after recovery had taken place. The pre-recovery filter had a 10 Hz cutoff for all coil channels except for the nose channels, which had a 5 Hz cutoff. Post-recovery filters had a 7.5 Hz cutoff.

4.1.3. Procedure

Participants had transducer coils placed on the "same" six flesh points as the model. The average participant tongue tip to tongue body distance was 4.1 cm (as compared to the model's 4.2 cm distance). Included in this average are tongue coil distances for subjects for whom acoustic analysis indicates imitation as indexed by their F2–F1 pattern.

E-A-RTONE 3A insert earphones (Aearo Company, Indianapolis, Indiana, USA) were used. They have a relatively flat frequency response between approximately 100 Hz and 4 kHz. Each earphone consisted of a piece of plastic tubing inserted through a foam earplug on one end and connected to a small amplifier on the other. The amplifiers and wires that carried the signal to the audio output channel were kept distant from the transmitter coils that generate the articulometer's magnetic field.

4.1.4. Design

The stimuli were presented in two blocks. Each block consisted of a random ordering of 24 [l], 24 [ɫ], 24 /r/ and 24 /w/ tokens. Because stimuli having vowel durations greater than 3500 ms were excluded, each of the 12 utterances was presented twice per block to provide an adequate number of stimuli.

The written instructions for Experiment 3 were almost identical to those of Experiments 1 and 2. In contrast to Experiments 1 and 2, however, trials in Experiment 3 were not self-paced. Therefore, a sentence was added to the written instructions for Experiment 3 asking participants to notify the experimenters if they wanted a break between trials. Verbal instructions regarding the self-reporting of errors remained the same. No examples of good shadowing were played to participants in Experiment 3, which renders any

[5] An error was categorized (loosely) as an end-tidal volume error if the participant did one or more of the following on a trial: yawned or exhaled sharply while shadowing, stopped producing the initial vowel and did not begin shadowing again on that trial, paused noticeably before producing the CV syllable, and/or did not produce a CV syllable or a final vowel.

finding of imitation here less attributable to the participant being primed to imitate.

Each participant's speech was recorded directly to hard disk at 20 kHz using the same shotgun microphone used to record the model stimuli. Although the microphone position was held constant across model and subjects, input gain was adjusted according to the vocal intensity pattern that each talker fell into during pre-experiment attempts at gain setting. All articulatory data were acquired at a 500 Hz sampling rate. Movement curves were filtered as described in the 'Stimuli' section above.

After data-collection, participants were asked to complete a debriefing form. This form explained the purposes of the experiment, and solicited participant assumptions about the purpose of the experiment and about the identity of the consonants (see Section 5.3, below).

4.1.5. *Articulator coil placement*

Henceforth, articulators are referred to by their initials, and whether they represent horizontal (X) or vertical (Y) movement in the occlusal bite plane. Articulators tracked included the tongue tip (TTX and TTY), the tongue body (TBX and TBY), the upper lip (ULX and ULY, with ULX taken as an index of lip protrusion, henceforth, LP), and the lower lip (LLX and LLY), though untransformed data on vertical displacement of the lower lip were not analyzed because it was not possible to partial out the contribution of mandibular movement to lower lip positions in the absence of jaw data. (We had originally planned to collect mandibular movement data for all participants, but the jaw data were not reliable for the model throughout the entire data-collection session, so we dispensed with the jaw for the participants, as well.)

The participant's head was oriented in the magnetic field in such a way that movement curves became more positive as they moved in an anterior and superior direction, and more negative as they moved in a posterior and inferior direction. However, because absolute vertical displacement of the tongue tip transducer does not always correspond to the tightest constriction degree, the tongue tip trajectory was rotated to the slope of the relevant section of the palate trace as in Honorof and Browman (1995) and Honorof (1999). This transformation yielded curves TTCL, or the constriction location along the palate, and TTCD, or constriction degree. These derived curves were used in place of TTX and TTY.

Because [l]s had a reduced tongue body gesture and [ɫ]s had a reduced tongue tip gesture, the two consonants did not share a measurable movement curve. Therefore, an algorithm was written in MATLAB (The Mathworks, Inc., Natick, Massachusetts, USA) to identify in the acoustics a landmark that would serve as a rough temporal marker of /l/ midpoint for both [l] and [ɫ] variants. Specifically, acoustic signals were filtered to increase the fidelity of MATLAB formant tracking using a 9th order high pass Butterworth filter with a 150 Hz cutoff (24 LPC coefficients, frame length = 320 samples [model], 400 samples [participants]; step size = 80 samples [model]; 100 samples [participants]). When there was no F1 value calculated for a given LPC window, the missing F1 value was replaced with the average of the nearest two non-missing F1 values. In the case of an odd number of missing values in sequence, the median was set to the average value of the two nearest neighboring non-missing values. In the case of an even number of missing values, the middle two were set to the average value of the two nearest neighboring non-missing values. Once the interpolation procedure had filled in all missing values by iterative application, the output was smoothed with a moving average filter (window size = 3, step size = 1).

For each token, a selection head was set during the initial [ɑ] F1 steady state prior to closure, and a selection tail was set during the final vowel after release. The minimum and maximum F1 value of the selected region was logged. Once all of the cases had been processed in this way, the absolute minimum was subtracted from the absolute maximum to yield the absolute range.

Fig. 1 depicts the labeling of /l/ schematically. Labels were placed at points during the closure and release transitions where F1 crossed a critical limit. The critical limit was computed by calculating the sum of a case's F1 minimum and 15% of its absolute range. This percentage was chosen because it resulted in F1 generally crossing the critical limit during the closure and release transitions for /l/, but not during the vowels.

The first accurate label placed by the algorithm was logged as /l/ closure, and the last accurate label was logged as /l/ release.

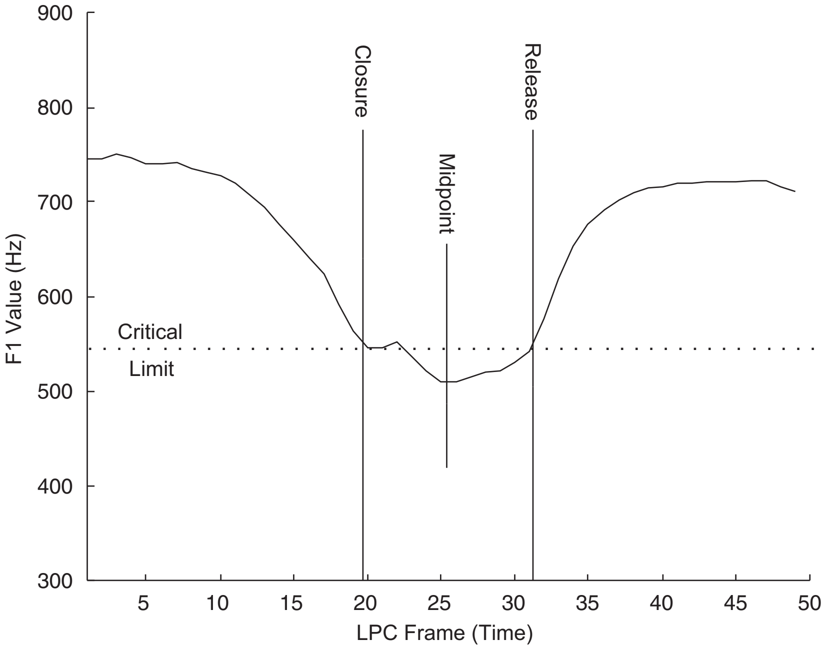

Fig. 1. Experiment 3: Schematic diagram of formant labeling.

When the algorithm failed to produce accurate labels for a generally accurate formant trace, labels were manually placed at points during the closure and/or release transitions where F1 seemed to come closest to the critical limit. Irrespective of how labels were placed, the temporal midpoint of the two labels was computed, and the X and Y articulator movement data at that midpoint were logged.

4.2. Results

4.2.1. Data exclusion

4.2.1.1. Acoustic data exclusion. Criteria for classifying errors and outliers in the acoustic data were identical to those described for Experiment 1. Using these criteria, fewer than 1% of the [l] and none of the [ɫ] data were removed due to participant error. In addition, 10.1% of the [l] and 9.2% of the [ɫ] data were removed because of difficulties encountered in applying LPC formant tracking used in calculating F1 minima. Fewer than 1% of the [l] and fewer than 1% of the [ɫ] data were removed due to data acquisition errors. Outliers accounted for 4.7% of the [l] and 3.7% of the [ɫ] data. Overall, approximately 16% of the [l]s and approximately 14% of the [ɫ]s were removed for one or another of the aforementioned reasons.

4.2.1.2. Articulatory data exclusion. Cases that qualified as participant errors in the acoustic analysis were removed from the articulatory analysis, accounting for fewer than 1% of the [l] and none of the [ɫ] data. Fewer than 1% of the [l] data and, fewer than 1% of the [ɫ] data were removed due to acquisition errors. In addition, 2.1% of the [l] and 5.8% of the [ɫ] data were removed due to difficulties encountered in LPC formant tracking used to establish the temporal midpoint of the lateral. For reasons introduced below, we also analyzed /w/ data, fewer than 1% of which were excluded due to acquisition errors.

Articulator positions 2.5 standard deviations greater than or less than the group mean were classified as outliers and removed from the dataset. Table 1 displays the percentage of outliers removed for each consonant-by-articulator combination.

4.2.2. Stimulus verification: [l] vs. [ɫ]

4.2.2.1. Model articulation: tongue and lips. The model attempted to produce laterals that de-emphasized one or another midline gesture, and that therefore retained, insofar as possible, a normally articulated primary gesture. Specifically, the model aimed to produce (a) [l] with less backing and less lowering of the tongue body into the oropharynx than typical English for his [l] and, conversely, (b) a reduced apical constriction for [ɫ]. In Fig. 2, we plot the model's articulation of these two /l/-types and, for comparison, /w/. The cross-token averages of tongue tip and tongue body coil positions that are plotted therein with respect to the palate trace suggest that the model was able to produce his intended articulations. The tongue tip gesture appears to be reduced for [ɫ]—perhaps even to zero—and his tongue body backing gesture seems to be reduced for [l]. By way of comparison, the model's tongue tip is not raised for [w], neither is his tongue body lowered into the oropharynx. Thus [w] looks to be canonically velar as one might expect, which suggests that 'pharyngeal' (rather than 'velar') might indeed be the most appropriate label for the darker lateral, though, as mentioned above, the rearmost coil is necessarily

Table 1
Percentage of articulator outliers (Exp. 3).

	[l] (%)	[ɫ] (%)	/w/ (%)
LP	1.1	2.1	1.6
TTCL	2.6	3.7	3.7
TTCD	3.1	2.1	1.1
TBX	< 1	2.6	2.1
TBY	4.2	3.1	2.6

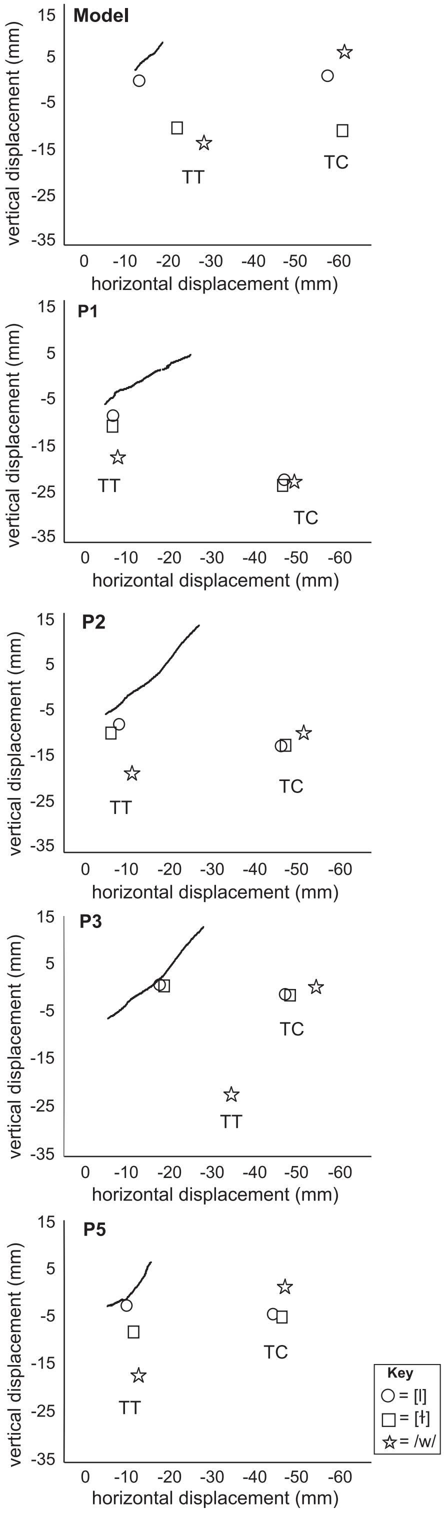

Fig. 2. Experiment 3: Two-dimensional displays of mean midsagittal articulator-coil positions at the temporal frame measured for the model and for those participants for whom acoustic analysis indicated dispositional imitation. Anterior segments of midline palate traces are provided for purposes of visual orientation. Palate angle is accurate, but a uniform minimal translation is applied to aid in visual interpretation. The scale along the abscissa applies to the tongue body (TB) coil positions. (Measurements for the tongue tip (TT) coil were based on a rotation with respect to a line fitted to the segment of the palate displayed here.)

somewhat distal to the actual oropharyngeal constriction location we infer so we cannot be certain of the exact position of the tongue dorsum and root in the pharynx.

Statistical tests were applied to these data and to lip data as well. Specifically, an ANOVA was run on the three lingual articulator-dimensions about which we have specific hypotheses, TTCD, TBX and TBY, as well as on TTCL and LP. As predicted, there were no significant differences in LP between the two types of /l/ ($p = .8011$). At a significance level of $p < .05$, results indicate that the model followed the instruction to reduce one gesture for each /l/ type. Specifically he succeeded in reducing the constriction degree of his tongue tip gesture for the [ɫ]. Conversely, he succeeded in reducing tongue body retraction and lowering into the oropharynx for [l] (TBX, TBY). Descriptive statistics (including differences between means) appear in Table 2a, with inferential statistics summarized in Table 2b.

Table 2a
Model articulator means and standard deviations (Exp. 3).

	/l/ type	*M* in mm (SD)	Light–dark
LP	**[l]**	9.85 (0.89)	−.09
	[ɫ]	9.94 (0.86)	
TTCL	**[l]**	−8.17 (1.11)	−.98
	[ɫ]	−7.19 (0.77)	
TTCD	**[l]**	−8.65 (0.61)	13.10
	[ɫ]	−21.75 (0.89)	
TBX	**[l]**	−53.47 (1.02)	5.04
	[ɫ]	−58.51 (1.14)	
TBY	**[l]**	0.11 (1.91)	11.99
	[ɫ]	−11.88 (1.78)	

Table 2b
/l/ model articulator ANOVA summary (Exp. 3).

	Source	SS	df	MS	*F*
LP	**Between**	.050	1	.050	.065††
	Within	16.737	22	.761	
	Total	16.787	23		
TTCL	**Between**	5.811	1	5.811	6.351†
	Within	20.128	22	.915	
	Total	25.939	23		
TTCD	**Between**	1030.643	1	1030.643	1784.094*
	Within	12.709	22	.578	
	Total	1043.352	23		
TBX	**Between**	152.763	1	152.763	130.147*
	Within	25.823	22	1.174	
	Total	178.586	23		
TBY	**Between**	862.441	1	862.441	253.004*
	Within	74.994	22	3.409	
	Total	937.434	23		

Note: For both conditions, $n = 12$.
Bonferonni adjusted significance level: $p < 0.01$.

* $p < .01$.
† $p = .019$.
†† $p = .801$.

4.2.2.2. Model acoustics: formant distance. For acoustic analysis of the model's utterances, the method for placing labels was identical to that used in Experiment 2, but only 20 LPC coefficients were used and the step size was reduced to 160 samples. An ANOVA was conducted on the model's F2–F1 data with '/l/ type' as the independent variable ([l] vs. [ɫ]). There were 12 [l] and 12 [ɫ] tokens included in the analysis. The main effect was significant ($F(1, 22) = 64.74$; $p < .0001$), with a greater mean F2–F1 value for the [l] ($M = 548.75$; $SD = 68.46$) than for the [ɫ] ($M = 349.17$; $SD = 51.92$), a mean difference of approximately 200 Hz. Although the model had been instructed to produce utterances for Experiment 3 as he did for Experiment 2, and while physiological data indicated that the model followed the instruction to reduce one midline constriction for each '/l/ type' in Experiment 3, the acoustics pattern in an unexpected way. That is, Experiment 3 mean formant distances are virtually identical to those in Experiment 1 (199.58 Hz versus 199.26 Hz), but mean formant distance is clearly not comparable between the latter two experiments; results of a two-way ANOVA on F2–F1 values revealed a significant interaction ($F(1,68) = 5.076$; $p = .03$) between factors '/l/ type' ([l] versus [ɫ]) and 'experiment' (2 versus 3), with [l] having a smaller mean formant distance in Experiment 3 than in Experiment 2, but [ɫ] having a larger mean formant distance in Experiment 3 than in Experiment 2. (The interaction notwithstanding, the direction of difference in mean distance was the same across all experiments, formants being further apart for [l] than for [ɫ].)

4.2.2.3. Summary. The articulatory data confirm that the model produced [ɫ] and [l] laterals with reduced tongue tip and tongue body gestures, respectively. They also confirm that the model's [ɫ] is apparently pharyngeal here, unlike his /w/ which is apparently velar in tongue body position. The model's laterals are unlike his /w/ in another way as well: they do not involve lip protrusion. Acoustic analysis (formant distance) is consistent with the articulatory analysis, though the size of the effect is relatively small.

4.2.3. Shadower acoustics: formant distance ([l] vs. [ɫ])

The same labeling method was used for acoustic analysis of the participants' disyllables as for analysis of the model's, but with 24 LPC coefficients and a slightly larger step size (200 samples). Participants' F2–F1 averages were entered with the same factor as the model ('/l/ type') into a repeated measures ANOVA ($N = 5$). The direction of the participants' [l]–[ɫ] mean difference was consistent with the direction of the model's mean difference. The main effect leaned toward significance ($F(1,4) = 6.89$, $p = .06$), with a greater F2–F1 mean value for the [l] ($M = 666.38$, $SD = 178.95$) than for the [ɫ] ($M = 533.94$, $SD = 237.03$), a difference of approximately 132 Hz—about six times the difference found in Experiment 1 and about twice that found in Experiment 2, but still a smaller distinction in formant distances than found for the model's target utterances.

Individual ANOVAs were conducted for each participant to determine whether they imitated the model. For all participants except P4, distance between the first and second formants was significantly larger for [l]s than for [ɫ]s as expected ($p < .05$; see Tables 3a and 3b). Significant mean differences in formant distance ranged from 87 Hz for P3 to 297 Hz for P5 and were always in a direction consistent with the model's mean difference in formant distance. For P4, because the tiny difference in mean formant distance between [l] and [ɫ] (less than 8 Hz) was statistically non-significant ($p = .62$), we conclude that P4 perceived no distinction between /l/ variants, or perceived the distinction but did not imitate it in any straightforward way that can be read off formant

Table 3a
/**l**/ F2–F1 means and standard deviations (Exp. 3).

	/l/ type	***n***	***M* in mm (SD)**
P1	**[l]**	45	599.49 (96.68)
	[ɫ]	46	423.26 (69.51)
P2	**[l]**	41	568.90 (126.63)
	[ɫ]	34	458.88 (122.02)
P3	**[l]**	46	985.00 (87.14)
	[ɫ]	43	898.37 (62.86)
P4	**[l]**	41	607.81 (61.30)
	[ɫ]	44	615.00 (70.73)
P5	**[l]**	31	570.68 (122.26)
	[ɫ]	42	274.19 (59.28)

Table 3b
/l/ F2–F1 ANOVA summary (Exp. 3).

	Source	**SS**	**df**	**MS**	***F***
P1	**Between**	706 445.842	1	706 445.842	100.01*
	Within	628 670.114	89	7063.709	
	Total	1 335 115.956	90		
P2	**Between**	224 980.807	1	224 980.807	14.50*
	Within	1 132 693.139	73	15 516.344	
	Total	1 357 673.947	74		
P3	**Between**	166 783.055	1	166 783.055	28.58*
	Within	507 638.047	87	5834.920	
	Total	674 421.101	88		
P4	**Between**	1098.737	1	1098.737	.25†
	Within	365 428.439	83	4402.752	
	Total	366 527.176	84		
P5	**Between**	1 567 831.078	1	1 567 831.078	187.89*
	Within	592 451.250	71	8344.384	
	Total	2 160 282.329	72		

* Bonferroni adjusted significant level: $p < .05$.
† $p = .62$.

distances. Therefore, tests on articulatory measures were run only for the remaining four participants (P1, P2, P3 and P5).

As we reported above, in Experiment 2 the model's mean formant distance was large—approximately 259 Hz. The magnitude of this distance is here interpreted to reflect the exaggerated distinction between /l/ variants intended by the model, however achieved. Here, in Experiment 3, physiological measures confirm that the model indeed achieved an exaggerated distinction between /l/ variants by adopting the intended gestural-reduction strategy. Therefore we predicted a similar pattern of mean formant distances in Experiments 2 and 3. However, the difference in mean formant distance in Experiment 3 was actually much closer to the mean formant distance between the fully bigestural /l/ variants of Experiment 1 than that between the [l] and [ɫ] laterals of Experiment 2. It may be that, without realizing it, the model achieved an exaggerated distinction between /l/ variants differently in Experiment 3, perhaps in compensation for the challenge of speaking with coils affixed to the tongue. This makes direct comparison of acoustic patterns across experiments potentially problematic.

4.2.4. *Model and imitator articulation: /w/ vs. /l/*

As our first measure of articulation, we examine evidence for or against lip protrusion for [ɫ]. Gesture-based theories of perception do not require a perfect fit between model utterances and imitated utterances, so a finding of lip activity for the imitators' [ɫ] in the absence of lip activity for the model's [ɫ] would not argue against direct realism or motor theories. Gesturalists would predict only that, if shadowers here imitate, they imitate tongue retraction and tip reduction in some small way, irrespective of whatever other behaviors may emerge. However, a finding of apparent acoustically or auditorily motivated enhancement would bolster acoustic or auditory theories of perception. Therefore, we consider acoustic evidence, then articulatory evidence, for a lip-protrusion enhancement strategy on the part of any participant who imitated.

4.2.4.1. Direct discriminant analysis. In designing our stimuli we assumed (following Wood, 1979) that /u/, and, by extension, /w/, is velar (that is, with a tongue body raised in the direction of the velum)—thus that [ɫ] with its tongue body retracted down into the oropharynx would not likely be confused with the labiovelar distractor /w/. Nevertheless, we ran direct discriminant analyses to assess the validity of our assumption.[6] Specifically, we asked whether the model protruded his lips for /w/ but not for either lateral while the imitators added lip protrusion to their constellation of gestures for the darker lateral to enhance a 'percept of darkness'. If not, that is, if our model *and* imitators did not protrude the lips, we must assume that the evidence for imitation as indexed by formant distances (reported above) must be attributable only to imitation of lingual articulation, not to labial substitution or lingual articulation plus labial enhancement.

The curves LP, TTCL, TTCD, TBX, and TBY were entered into each talker's analysis at once. These analyses allow us to determine the number of dimensions along which our three sounds reliably differ for each talker. For three groups ([l] versus [ɫ] versus /w/), two discriminant functions are extracted. These functions define two orthogonal linear hyperplanes that optimally separate the data into three disjoint classes such that prediction error is minimized across classes and variance is maximally dispersed. Here, three statistics are relevant to the interpretation of our results: dispersion of group centroids, correlations of group membership with our set of articulator-dimensions (henceforth, *predictors*), and classification of individual cases. Results for each of these steps follows.

4.2.4.1.1. Group centroids. We calculated a group centroid for each group-function combination, a centroid being the mean discriminant score for a group on a given function in output space. Relative dispersion of the centroids helps us determine how our three groups (/w/, [l] and [ɫ]) are separated by the function. For the model and each imitator, the separation of centroid values for at least the first function confirms that /w/ is discriminated from the /l/ categories. In Table 4, we report group centroid values for all three groups for the first function. For the model and all participants, either first-function centroids for /w/ are much farther from

[6] In direct (that is, standard) discriminant analysis, one enters all predictors into the analysis simultaneously. Shared variance among predictors contributes globally to the functions, but not to any particular predictor. The entry of all predictors at once distinguishes direct discriminant analysis from hierarchical (that is, sequential) discriminant analysis where the order in which predictors enter the equations would be specified by the researcher. Hierarchical discriminant analysis is not called for in the present study because we have no obvious basis for setting a priority order among predictors. In such cases one sometimes relies on stepwise discriminant analysis to assign some predictors higher priority order for entry into the equations on the basis of statistical criteria, but we did not do this here because we have no reason to require a reduced set of predictors. See Tabachnick and Fidell (1989).

Table 4
Discriminant analysis: group centroids Function 1, all predictors (Exp. 3).

	% of variance	Category	Group centroid
Model	64.8	**[l]**	12.756
		[ɫ]	−3.951
		/w/	−8.805
P1	97.4	**[l]**	2.178
		[ɫ]	1.761
		/w/	−4.053
P2	79.8	**[l]**	−.856
		[ɫ]	−.887
		/w/	1.583
P3	100	**[l]**	7.889
		[ɫ]	7.464
		/w/	−13.978
P5	61.2	**[l]**	−2.424
		[ɫ]	−.054
		/w/	2.598

/l/ centroids than /l/ centroids are from each other (model, P1, P2, P3), or all three centroids are roughly equally spaced (P5). The significance of this pattern is examined through correlations reported below. The present analysis was run in order to evaluate the potential confusion of /w/ and /l/, not in order to test discrimination of /l/ types, so no further results are reported (e.g., second discriminant function centroids).

4.2.4.1.2. Chi-square and percentage of variance. We employed direct discriminant analyses to explore the reliability of association strength between our set of predictors and group membership. For the model and for each of the four imitators, under direct discriminant analysis, chi-square indicates that /w/ centroids are reliably separated from /l/ centroids ($p<.001$; df=10). Percentages of variance accounted for by the first function range from 61.2% to 100% (see Table 4).

4.2.4.1.3. Classification matrices. We derived linear equations that classify cases into groups. Doing so allowed us to check the adequacy of classification, that is, to determine the ratio of cases correctly classified. We used a jackknifed design, excluding each case from the computation of the coefficients used to assign that case to a group. The resulting classification indicated that, for P3, only 10.5% of the [ɫ]s were misclassified as /w/. For P5, only 4.7% of the [ɫ]s were misclassified as /w/. For the model and the other two participants, none of the [ɫ]s were misclassified as /w/.

4.2.4.1.4. Summary of the discriminant analyses. Results of the three statistics reported for the discriminant analyses are unambiguous: /w/ is discriminated from both laterals reliably. This suggests that neither the model nor the imitators were simply substituting /w/ for either type of lateral. This frees us to limit focus to the two types of lateral to the exclusion of /w/. However, detailed interpretation of loading matrices and contrasts in discriminant analysis is difficult and potentially controversial. Therefore, rather than running new discriminant analyses without /w/, we ran linear regression on the articulatory data with /w/ removed to determine which predictors are most helpful in separating the two types of lateral. Results from regression analyses follow.

4.2.5. Articulation: [l] vs. [ɫ]

4.2.5.1. Model and imitator—all predictors. We ran linear regressions on the relevant lip and tongue coil positions (LP, TTCD, TBX and TBY) for the model and for each imitator to determine which predictors were significantly correlated with each of the two types of /l/, and we verified that all of the predictors we included contributed significantly. Then, we ran a linear regression for each imitator using only those predictors shown to be significant via *t*-tests for that imitator in the first regression. No /w/s were included in the linear regressions.

Tests of goodness of fit of the regression equations for all four imitators were significant ($p<.05$). A summary of variances and correlations appears in Table 5a. Differences between [l] and [ɫ] means for each significant predictor appear in Table 5b along with additional statistics from the linear regressions. For all subjects, at least one predictor about which we have a specific hypothesis (TTCD, TBX or TBY) showed a significant linear correlation. In each of these cases, the pattern, that is, the direction of mean difference was the same for the imitators as it was for the model (cf. Table 2a). LP, a constriction about which our theory leads to no predictions, achieved significance for P2 and P5 only, but showed slight lip retraction (~1 mm) for [ɫ], not protrusion, for P2. This leaves only P5 as a possible 'labializer', but a labial enhancer at most; even for P5 the discriminant analyses revealed that only 4.7% of the [ɫ]s were misclassified as /w/. Furthermore, on debriefing, P5 reported having heard a type of /o/ with the "tongue tip not touching the top of [the] mouth."

This similarity of patterning in articulator positions between model and talkers can be seen in the mean plots in Fig. 2. The plot reveals that, at the time frame measured, the coil affixed to each participant's tongue tip (TT) was closer to the palate on average for the [l] than for the [ɫ]. Furthermore, the (mean) relative lowering and/or retraction of the coil affixed to the tongue body (TB) for [ɫ] is also evident, though the effect is smaller. Both forms of gestural reduction are seen in the plots of the model's productions. The mean for /w/ is also plotted for reference and is clearly separate from the two other means for all talkers.

While the regression was significant for all four participants, the equation for P3 accounted for a relatively small percentage of the variance (adjusted $r^2=.045$). In this connection, we note that, during debriefing, P3 reported having initially assumed that the model was "making a mistake" while producing [ɫ]s, and that, she, therefore, started out intentionally not imitating the [ɫ]. An overall imitation effect nevertheless emerges in the regression.

4.2.5.2. P2 and P5 lingual predictors only. The ANOVAs whose results are reported under Section 4.2.2.1 confirm that there were no significant differences in lip protrusion between /l/ types for the model. The model's native accent of English is not reputed to employ active protrusion of the lips for any /l/ variant. Neither was the model instructed to add a labial component to the target laterals. In fact, lip data had been collected initially just in case we decided to analyze the fillers /r/ and /w/, an analysis which we did not, in the end, see a point in doing for /r/. We ultimately wish to know whether participants were imitating (inherently non-contrastive) gestural differences between /l/ types irrespective of whatever else they may have been doing with their lips to augment the difference. Our theoretical model does not lead to the prediction that there should be no significant differences between lip

Table 5a
Summary of linear regression (all predictors, Exp. 3).

	df	*F*	r	r^2 (adj.)
P1	***2,85***	17.778*	.543	.278
P2	***3,78***	12.800*	.574	.304
P3	***1,89***	4.189*	.212	.045
P5	***3,84***	96.602*	.881	.767

* $p<.05$.

Table 5b
Linear regression and means (all predictors, Exp. 3).

	Predictor	Beta	*t*	*M* (SD)		Light–dark
				[l]	[ɫ]	(*M* direction)
P1	**TTCD**	−.448	−4.702*	−9.665 (.920)	−10.879 (1.098)	1.214 (o)
	TBY	−.202	−2.125*	−23.649 (1.608)	−24.860 (1.752)	1.211 (o)
P2	**ULX**	−.707	−4.806*	7.472 (1.090)	7.356 (.782)	.116
	TTCL	.553	5.311*	−3.258 (3.907)	−.844 (2.954)	−2.414 (o)
	TBX	−.640	−4.640*	−46.670 (3.774)	−47.646 (2.725)	.976 (o)
P3	**TBX**	−.212	−2.047*	−45.160 (1.530)	−45.830 (1.594)	.670 (o)
P5	**ULX**	.664	8.025*	8.280 (1.883)	11.986 (.989)	−3.706 (o)
	TBX	−.249	−3.003*	−44.441 (1.173)	−46.328 (1.042)	1.887 (o)
	TBY	−.271	−5.187*	−6.285 (1.968)	−7.078 (2.325)	.793 (o)

* $p < .05$; (o) same direction as model.

movements of the model and imitators. There may be such differences. In any case, the regressions reported in Section 4.2.5.1 certainly indicate lingual imitation irrespective of lip activity.

Given no prediction regarding the LP predictor, it occurred to us that a significant contribution of lip activity to the regression for two imitators may have cloaked the relative contribution of tip and body predictors for those participants (and indeed may have confused matters considerably for P2 whose mean LP direction indicated slight lip retraction for [ɫ] rather than protrusion). Therefore, we ran a second linear regression, this time entering only the lingual predictors. Such a regression was run for each participant who exhibited a significant difference in LP in the first regression (P2 and P5). Doing so allowed us to determine which lingual predictors were significantly correlated with each of the two types of /l/. Again, no /w/s were included. (Linear regressions were already run without LP for P1 and P3; see Tables 5a and 5b.)

Tests of goodness of fit of the regression equation for both participants were significant ($p < .05$). A summary of variances and correlations appears in Table 6a. Differences between [l] and [ɫ] means for each significant predictor appear in Table 6b along with additional statistics from the linear regressions. For both subjects, at least two of the predictors about which we have a specific hypothesis (TTCD, TBX or TBY) showed a significant linear correlation. In each of these cases, the pattern, that is, the direction of mean difference, was the same for the participants as it was for the model.

4.2.6. *Labial substitution/enhancement revisited—articulation: /w/ vs. [ɫ]*

4.2.6.1. Model and imitators: the lips alone. While we have no hypothesis regarding whether lip rounding might be used by imitators to enhance acoustic output in a way that makes tongue-body retracted [ɫ] 'sound' darker, such a finding might be predicted by theories that treat acoustic patterns (or representations of acoustic patterns) as the primitives of speech perception. In other words, an acoustic/auditory theorist might predict that at least some speakers some of the time would misattribute the smaller F2–F1 distances arising out of the model's tongue body backing gesture to lip rounding, and would therefore show more lip movement in at least some of their own shadowed responses to the model's own [ɫ]s in the absence of lingual imitation.

In fact, our reading of at least the seminal literature in the field (e.g., Chiba & Kajiyama, 1941) informs us that lip rounding lowers F2 and F1, not just F2—a trade-off that would not provide a basis for substitution of lip activity for tongue backing. However, to err on the side of competing theories, we acknowledge that it remains possible that some imitators might attend only to F2 lowering tied to tongue body activity, not to the distance between the lowest two formants, so it remains worth considering lip activity in connection with a competing theory according to which multiple articulatory strategies can be traded off in achieving an acoustic or auditory target. To this end, we ran direct discriminant analyses to find out whether shadowed productions of /w/ were discriminated from [ɫ] *on the basis of lip positions alone*. The curves LP and Lip Aperture (LA, the 2D Euclidean distance between ULY and LLY) were entered into each participant's analysis at once. For purposes of comparison, a direct discriminant analysis was also run on the model's articulation. For two groups, one discriminant function is extracted that optimally separates the data into two disjoint classes such that prediction error is minimized between classes, and variance is maximally dispersed. As before, we calculate the mean discriminant score (group centroid) for each group on the function in output space.

Table 6a
Summary of linear regression with only lingual predictors (Exp. 3).

	df	*F*	*r*	r^2 (adj.)
P2	***3,78***	7.470*	.472	.193
P5	***2,86***	59.690*	.762	.572

* $p < .05$.

Relative dispersion of the /w/ and [ɫ] groups indicate that they are well separated by the function for the model and for each participant. In Table 7, we report group centroid values for both groups. The significance of this pattern is examined through correlations in which Wilks' Lambda confirms the reliability of association strength between our set of predictors (LP and LA) and group membership; /w/ centroids are reliably separated from [ɫ] centroids ($p < .001$; df=2, for the model and each of the four participants). In all cases, means indicate that /w/ is produced with a smaller lip aperture than [ɫ]. For the model and P1, P2 and P3, /w/ is also slightly more protruded (< 2 mm). For P5, LP is actually slightly less for /w/ than for [ɫ] (~ 1 mm), but the canonical correlation for this participant is only a moderate .693 (versus a strong correlation of between .857 and .971 for the model and all other imitators).

Taken as a whole, this latter set of discriminant analyses for the lips show a consistent pattern where lip aperture is concerned; namely, clear discrimination of /w/ and [ɫ] lip positions with

Table 6b
Linear regression and means (only lingual predictors, Exp. 3).

	Predictor	Beta	*t*	*M* (SD)		Light–dark
				[l]	[ɫ]	(*M* direction)
P2	**TTCL**	.519	4.377*	−3.258 (3.907)	−.844 (2.954)	−2.414 (o)
	TTCD	.632	3.035*	−12.789 (6.755)	−13.565 (5.957)	.776 (o)
	TBX	−.662	−3.373*	−46.670 (3.774)	−47.646 (2.725)	.976 (o)
P5	**TBX**	−.750	−10.644*	−44.441 (1.173)	−46.328 (1.042)	1.887 (o)
	TBY	−.277	−3.927*	−6.285 (1.968)	−7.078 (2.325)	.793 (o)

* *p* < .05; (o) same direction as model.

Table 7
Direct discriminant analysis: group centroids for Function 1 for measures LP (ULX) and LA (Exp. 3).

	% of variance	Category	Group centroid
Model	100	[ɫ]	2.079
		/w/	−2.079
P1	100	[ɫ]	2.474
		/w/	−2.309
P2	100	[ɫ]	−1.840
		/w/	1.472
P3	100	[ɫ]	4.104
		/w/	−3.922
P5	100	[ɫ]	0.993
		/w/	−0.908

greater lip aperture for [ɫ] than for /w/ for the model and all imitators. The tests also show a largely consistent pattern where lip protrusion is concerned: [ɫ] and /w/ are distinct (that is, no talker is substituting /w/ for [ɫ] wholesale), and only one talker (P5) shows a possible trend in the direction predicted by acoustic/auditory theories that inspired this test of lip positions, but the evidence for lip protrusion for [ɫ] in that case is not strong.

4.3. Discussion

Experiment 3 provided acoustic data consistent with those of the first two experiments; all but one participant in the present experiment imitated the model's speech in the shadowing task without being instructed to imitate. Although previous work reveals a general disposition for talkers to imitate a model, it also indicates that they tend to undershoot model targets (Fowler et al., 2003; Sancier & Fowler, 1997; Shockley et al., 2004). It is not, therefore, surprising that three of the four Experiment 3 participants who exhibited significant imitative behavior according to our acoustic measure nevertheless produced a pattern of formant-distance-difference undershoot with respect to the model's F2–F1 means. We note, however, that lip protrusion was not seen in the productions of any participant who undershot the model's articulations along our acoustic measure. That is, those who 'under-imitated' the model's distinction in tongue shape between [l] and [ɫ] did not compensate by protruding the lips. Rather, they simply produced a less distinct distinction as reflected in formant distances. The only lip protruder, P5, actually exaggerated the model's formant distance differences between /l/ types (though P5 did not produce a lip closure gesture). Upper lip protrusion for P5 was on the order of approximately 3.7 mm greater for [ɫ] than for [l]. A mean difference of 3.7 mm is the largest difference seen for a significant predictor in the linear regressions and is not insubstantial for the upper lip. However, as with all the effects that emerge from the linear regressions, P5's lip protrusion effect is not very large compared with targeted movements of the model's articulators. Subsequent direct discriminant analysis of lip data comparing only /w/ and [ɫ] confirmed protrusion; lip protrusion is actually larger by 1 mm for [ɫ] than for /w/, but the canonical correlation was not strong.

We attribute the small size of the effects in Experiment 3 to the likelihood that our design elicits competing strategies within a single talker. That is, the instruction to shadow speech, while clear, may have triggered both the disposition to imitate and the disposition to rely on highly practiced speech motor routines. The remarkable fact is that, even in the absence of an explicit process of imitation, shadowers subtly imitated the model's gestures even though successful imitation of those gestures may have induced some measure of frustration.

If the model had really produced acoustic signals that were ambiguous with respect to the underlying gestures that structured them, one would imagine that some talkers would have recovered the set of gestures actually used in the (hypothetically) ambiguous acoustic signal and thus have reproduced those that they recovered, while others would have recovered and reproduced a set not used. Had this been the case, we would have expected to see imitation of the model's lingual gestures by only some speakers. In the magnetometer study, however, all speakers whose productions showed acoustic evidence of imitation also imitated at least some of the model's lingual gestures. Perhaps a larger sample might have produced such a 'substituter'. Certainly, on the basis of the present findings, at least, we observe that shadowers did imitate speech gestures dispositionally (that is, without having received explicit instruction to imitate) as seen in the results of the linear regression, but with participant-to-participant variability in which specific gestures were imitated best. We do not infer, however, that those gestures that were not imitated well by a given talker were necessarily not perceived. It may be that some forms of gestural organization, while perceptible, are not consistently well imitated by the average talker without practice or perhaps even without explicit articulatory training.

Cross-shadower variability exposes a pattern in the present data. The diversity of behaviors within and across experiments can be seen in the acoustic measures plotted by participant in Fig. 3. Plotted is the difference for each participant between mean [l] formant distance and mean [ɫ] formant distance. Plotted differences for those who had 'darker' [l]s than [ɫ]s fall below the solid horizontal line. These participants are considered non-imitators on the formant-distance measure. Those who imitate (that is, show a positive difference and thus a greater spread between F1 and F2 for [l]s) but fall short of achieving the model's mean difference in formant distance (the dotted line) are considered weak imitators. We draw particular attention, however, to the topmost data point in Experiment 3, the only participant whose mean formant distance difference indicates acoustic 'overshoot' (nearly 100 Hz greater

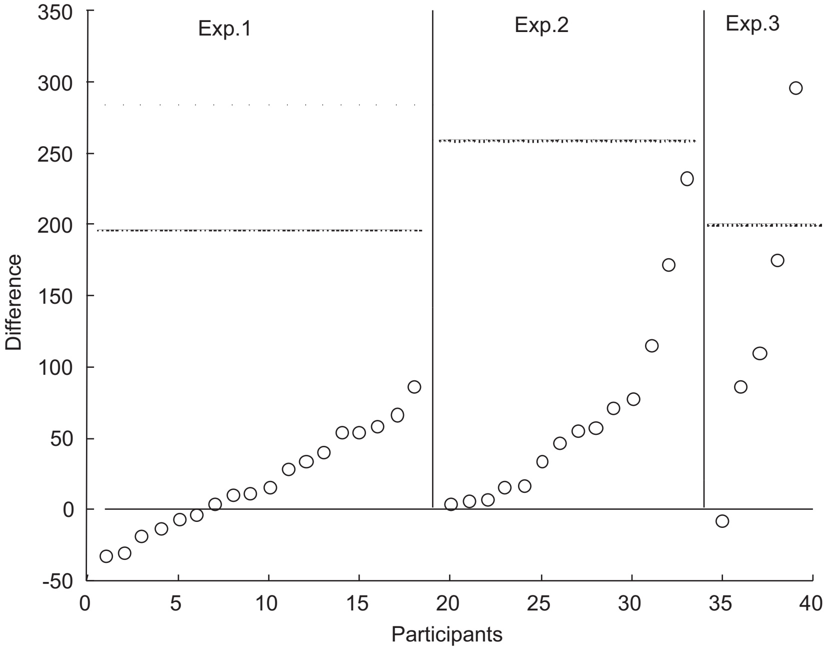

Fig. 3. Acoustic differences between /l/ types across experiments. The dotted line represented the model mean formant distance between [l] and [ɫ]. Exp. 1: Most imitate the model but only slightly. The six participants plotted under the solid line exhibit a pattern of formant distances opposite to that exhibited by the model, and thus are considered non-imitators. Exp. 2: All imitate the model, but most not strongly. Exp. 3: One does not imitate. One hyper-enhances chromatic differences with respect to the model's mean formant-distance difference. The remainder show slight differences in formant distance differences less distinct than the model's differences.

than the model's). This is P5, also the only participant whose /w/, [l], and [ɫ] group centroids (all predictors) were distinct as opposed to only /w/ versus /l/ being distinct, the only participant who used lip protrusion for [ɫ] to the point that approximately 4.7% of the [ɫ]s were misclassified as /w/ in the direct discriminant analysis and the only participant who exhibited a significant difference between /l/ types in both TBX and TBY. Ironically, given P5's formant-distance overshoot and relative imitative fidelity using the tongue, P5 was arguably the participant who least needed to use the lips to make [l] and [ɫ] distinct. If anything, P5's lip protrusion for [ɫ] may best be described as an enhancing rather than a compensatory strategy.

One might expect a talented imitator to attend more closely to an unusually dark [ɫ] than to an unusually light [l]; [ɫ] does not normally occur in (potentially) ambisyllabic position, and our [ɫ] was also strangely 'dark-sounding' to the investigators' ears. Indeed, upon debriefing, all four imitators in Experiment 3 reported having heard an /l/ that they described as "blechy-yucky", "weird", "swallowed" or "unfamiliar". P3 even reported hearing the tongue not touching the "top of the mouth." No one reported hearing non-English sounds, neither did anyone report not hearing speech sounds at all—which is not surprising because they were primed to hear the speech as speech by the instructions which indicated syllable affiliation of the target sound as well. Simply put, it may be that strong imitators fixate on the unusual and different and exaggerate it. Overshoot of whatever is remarkable is also consistent with published findings on patterns of imitative behavior (caricature) in professional impersonators (Zetterholm, 1997).

As already noted, overall, effects were small. Participants' [ɫ]s in all experiments were generally not like prototypical coda [ɫ]s. Participants were not instructed to imitate, and thus are not, in all likelihood, imitating on purpose. Their own speech habits must have been competing with any tendency to imitate gestures. However, any imitation of unrehearsed gestures, we maintain, implies perceiving them, and all who imitated, imitated at least one gestural difference between the model's unusual liquids.

5. General discussion

5.1. Overview

Taken together, results of all three rapid shadowing experiments lead us to a single conclusion: Despite individual differences in disposition to imitate and in fidelity of the match, when speakers do imitate, they reproduce aspects of the model's articulation even when the sound so produced in a given syllable position or context is in some way unrepresentative of the imitator's own phonology.

Acoustic evidence (F2–F1 formant distances) from the first experiment in which position was controlled indicates that those who imitated reproduced distinctions between [l] and [ɫ]. However, the degree of imitation was small: model acoustics were undershot by approximately 100–200 Hz by all imitators.

In an attempt to make the difference between /l/ variants greater, we ran a second experiment in which the model produced similar utterances in a more unnatural way. Specifically, he produced especially 'light' and especially 'dark' laterals by reducing the magnitude of the tongue body or tongue tip gesture, respectively, again with position controlled. All participants imitated these positionally unpredictable differences, and some came closer to matching the extent of the difference for the two /l/ variants in terms of the model's mean formants distances than did the participants of Experiment 1.

A third experiment was designed to directly investigate articulation in a task comparable to that of Experiment 2. Our hypothesis was that imitation of /l/ type would be indicated in the acoustic record again, and that, when it was, gestural imitation would also be evident. We reasoned that such gestural imitation, if it obtained, must indicate that imitators succeeded in perceiving positionally unpredictable aspects of the model's articulation. Although the acoustics of only four of five participants of Experiment 3 suggested imitation, of those four, three produced acoustic differences between /l/ types that were fairly close to the model's acoustic differences and one actually produced a difference in formant distance greater than the model. Discriminant analyses were run on the articulatory data of these four participants. Results clearly revealed different midline tongue shapes for [l] versus [ɫ]. Linear regressions were then run to help determine which articulatory dimensions were relied upon most heavily in the imitation. For all subjects, at least one predictor about which we have a specific hypothesis (TTCD, TBX or TBY) showed a significant linear correlation. In each of these cases, the effects are small, though significant and in the predicted directions where predictions were made (that is, the imitators' articulation reproduced key aspects of the model's articulation).

One criticism of Experiment 3 might be that we have examined tongue body data that are taken from a flesh point somewhat anterior of the true tongue dorsum, and certainly far anterior of the tongue root. We recognize that we lack information about the exact positions of true dorsal and pharyngeal flesh points. Nevertheless, it is clear that retraction of the more posterior surfaces of the tongue body as a whole must be reflected in the backing of the tongue body coil, so we believe that we are safe in using our measurements as an index of tongue-body retraction. Comparison of tongue body coil positions for /w/ and [ɫ] confirmed that the constriction for the lateral is lower and more posterior, and thus not consistent with a possible velar target along the midline. Regardless, the exact constriction location of the tongue body gesture for [ɫ] is not relevant to our claim that imitators lowered and retracted the tongue body and/or reduced the constriction degree of the tongue tip gesture for [ɫ]. Minimally, it is clear that in no case did imitators simply substitute their own normal [l] and [ɫ] for the model's exaggeratedly distinct chromatics of the laterals in Experiment 3.

We know that people are capable of shadowing very quickly indeed (e.g., Fowler et al., 2003). This finding is interpreted as evidence of rapid perceptual access to the gestures needed for imitation. Our shadowers were probably perceiving in the same way here. The exact nature of the perception and imitation may have differed subtly from real world perception and gestural imitation due to the constraints necessarily imposed by the design. In this connection, an anonymous reviewer suggested that each of the three experiments discussed here employed a 4-choice reaction time paradigm which may have biased responses in an unknown way. That is, on every trial, the subjects may have "pre-activated" four gestural constellations (/r/, /w/, [l] and [ɫ]), which may have influenced production. For instance, the gestural constellations may have become blended and less distinct or, alternatively, the gestural constellations may have become more distinct due to inhibitory mechanisms. Certainly, however, irrespective of the details of when, how and why gestures are perceived, reproduction of gestures seen here fully implies gestural perception.

5.2. Theoretical accounts of the findings

We interpret our findings within the context of a direct realist theory of speech perception. In that theoretical account, listeners to speech extract acoustic information about gestures and use that information to perceive speech gestures. Following complementary claims made within Articulatory Phonology (e.g., Browman & Goldstein, 1992), speech gestures are defined as phonological speech actions, and, at the same time, phonetic actions (see Benus & Gafos, 2007). In this regard, perceiving speech is like perceiving generally (e.g., Fowler, 1986, 1996). That is, in all instances, given proximal stimulation at the sense organs, perceivers extract information about the distal sources—those objects and events in the environment that structured the information. Perceivers use that information to perceive their environment. The very short latencies that speakers can demonstrate when they shadow speech in a choice reaction time setting (e.g., Fowler et al., 2003) occur, in this account, because of the extreme compatibility between perceived stimuli and required responses. The speech stimulus and response are both gestural. The present findings that gestures are imitated in the shadowing task further support a gestural account.

Except for the motor theory of speech perception (see Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985), in our view, no other account of speech perception apart from direct realism, in its present formulation, handles either the shadowing findings of Fowler et al. (2003) or the present findings of gestural imitation. Motor theorists propose not only that gestures are perceived, but also that listeners recruit their own speech motor systems in the course of perceiving speech. There is now evidence for this (e.g., Fadiga, Craighero, Buccino, & Rizzolatti, 2002) including evidence that selective potentiation of the speech motor system using transcranial magnetic stimulation speeds speech perception in a correspondingly selective way (D'Ausillo et al., 2009). In the motor theory, short shadowing latencies are possible because perceiving speech primes the motor system to produce what has been perceived. In the present research, gestures are imitated for the same reason.

Accounts of speech perception in which immediate perceptual objects are auditory/acoustic (e.g., Diehl et al., 2004) do not, without elaboration, predict rapid shadowing or gestural imitation at all. However, theorists in this domain, among others, have posed a challenge for a direct realist account. Specifically, they argue that the inverse mapping from acoustic signals to gestures is one-to-many and hence is indeterminate. Accordingly, an account of the present findings and many others in terms of gesture perception can be ruled out.

Specifically, it has been claimed that more than one possible gestural configuration across or within phonological contexts can produce a single acoustic pattern, whether the articulatory variability be from dialect to dialect or from vocal tract configuration to vocal tract configuration. (See, for example, Atal, Chang, Mathews, & Tukey, 1978; Delattre & Freeman, 1968; Guenther et al., 1999; Lindblom, Lubker, & Gay, 1979; Riordan, 1977; Sondhi, 1979.) Proponents of one or another version of the *many-to-one-mapping hypothesis* usually argue that articulatory variants are not associated with salient perceptual differences, thus that purported acoustic or auditory stabilities are the objects of perception. Within such a model, the listener (or speech recognition algorithm) must pass through one level or more of translation, interpretation, interpolation or filtering to recover vocal tract shapes from the waveform. (For a review of the issues from a non-gesturalist perspective, see Diehl et al., 2004.) If such theories were correct in asserting that listeners recover acoustic stabilities, not underlying gestural configurations, it would follow that, in the present research, rapid shadowers presented with gestural configurations that did not match the gestural configurations of their own linguistic codes should have failed to consistently reproduce them. Given a large enough sample, an acoustic target model would predict a greater variety of articulatory strategies (for instance, stronger evidence of compensatory lip activity) than we found, and would attribute such variety to acoustic pattern-matching on the part of the shadower.

The belief, widely held among speech researchers, that there is a many-to-one relation between raw articulatory speech detail and acoustic detail leads to the supposition that precise vocal tract area functions could not be recovered from the acoustic output even if all the acoustic poles and zeroes were known up to infinite frequencies. (For discussions of some relevant issues, see Borg, 1946; Gopinath & Sondhi, 1970; Kac, 1966.) Many have therefore given up on inversion and instead sought for invariance in acoustic properties (again, see Diehl et al., 2004). Others have continued to look for invariances in articulation, and have sought to develop more advanced recovery techniques (e.g., Hogden et al. (1996); Hogden, Valdez, Katagiri, & McDermott, 2003; Yehia, 1997). We believe the entire problem is misconceived. It seems to us that listeners need neither an inversion strategy nor a computational mapping process mediated by acoustics; listeners do not actually need to recover precise vocal tract area functions from the time-varying signal. Rather, for purposes of basic decoding of the talker's phonologically encoded message, the listener primarily needs to perceive just the information that broadly specifies temporally overlapped, linguistically relevant events in the vocal tract irrespective of the type and amount of linguistic, paralinguistic, sociolinguistic, and nonlinguistic information in the time-varying speech signal that the listener actually perceives and stores. Many of the details of vocal tract shape that may be difficult or impossible to recover from acoustics directly would have to be considered noise from the linguist's perspective. Fortunately, Heinz (1967), Mermelstein (1967) and Schroeder (1967) have shown that it is possible to recover gross vocal tract shapes sufficient to determine linguistic category membership. Although a mapping is implicated in this work, it is a one-to-one mapping between a limited class of acoustics and the minimally specified vocal tract shapes that produced them. Much work remains to be done on articulatory recovery of temporally co-produced gestures, but there is no reason to complicate matters with a search for acoustic invariances that require additional layers of processing.

Within another family of speech perception theories (e.g., Goldinger, 1998; Johnson, 1997a; Palmeri, Goldinger, & Pisoni, 1993), perception of speech is held to result in storage of episodic traces. These traces code phonetic properties of utterances and nonlinguistic properties, such as information about the speaker's voice. Such episodic accounts address a set of findings that speech perceivers do not "normalize" speech in the sense of stripping off and discarding nonlinguistic information in the course of phonetic perception. Perhaps, it is argued, the mental lexicon is a collection of episodic traces that preserves phonetic and nonlinguistic detail about speech events. The account attempts to explain why, when imitation occurs, that imitation might be of subphonemic properties of speech as we found in the present research. However, the account does not predict that imitation must occur. We do not dispute evidence for preservation in memory of phonetic and nonlinguistic detail about speech events, although we interpret any preserved "detail" that was produced by the speaking vocal tract as articulatory in nature. However, even though evidence for the idea of an episodic speech memory was obtained, in part, by observing imitation in speech listeners (Goldinger, 1998), imitation was not a prediction of the episodic theory. Rather, Goldinger made use of a previous finding that listeners *do* imitate to motivate the design of his research.

Mitterer and Ernestus (2008) have offered a different challenge to a direct realist account of rapid shadowing and imitation findings. They proposed that speech perceptual objects that underlie perception-based production are neither auditory/acoustic nor episodic, but are abstract and phonological. Using a shadowing task somewhat different from that of Fowler et al. (2003), they found little evidence for Dutch speakers' imitation of a model speaker's production of an alveolar or uvular trill, with speakers tending to stick to their own preferred place of articulation. They also found that shadowing latencies were not slower when the model's gestures for producing the trill mismatched those of the shadower. Finally, they concluded that phonetic detail is only imitated if it is phonologically relevant. They attributed the finding by Fowler et al. (2003) that shadowers extended voiceless stop VOTs when those of a model speaker were lengthened to phonological relevance. They cited the fact that, in English, there are both aspirated and unaspirated allophonic variants of voiceless stops.

Their first observation, of limited imitation overall, is not wholly incompatible with the finding of Fowler et al. (2003). The latter investigators found that, although speakers did imitate the VOTs of voiceless stops in the sense that they extended VOTs when the model's VOTs were lengthened, the extensions were much smaller (8 ms in one experiment, 4 ms in another) than the lengthening (57 ms). Shadowers largely maintained their habitual way of talking. That, in many cases, participants failed to imitate altogether in the study of Mitterer and Ernestus (2008) but not in that of Fowler et al. (2003) or in the present research, may be due to differences in the shadowing procedure in the former study. In particular, Mitterer and Ernerstus had participants shadow a pair of syllables separated by 500 ms, rather than shadowing a disyllable as in Fowler et al. (2003) and in the present study. This may be why latencies were much longer in the former study than in the study of Fowler et al. It stands to reason that a longer interval between perception of a syllable and its production can induce one to forget detail. Indeed, such a claim has been invoked in interpretation of findings in categorical speech perception studies (e.g., Pisoni & Lazarus, 1974; see also Frankish, 2008).

Mitterer and Ernestus (2008) found that shadowing latencies were unaffected when model and participant gestures were mismatched. This finding is incompatible with findings of Fowler et al. (2003). The latter study compared simple response latencies on trials on which place of articulation of the model's and shadowers' syllables matched and mismatched. There was a significant latency advantage on matching trials. The simple reaction times in the Fowler et al. study were very fast in comparison with the substantially slower latencies obtained by Mitterer and Ernestus, which may explain the differing results.

Mitterer and Ernestus' (2008) also suggested that phonetic detail is imitated only when it is phonologically relevant. This suggestion is also contradicted by the findings of Fowler et al. (2003). In Experiment 4 of the latter study, the model's voiceless VOTs were either lengthened (averaging 130 ms) or not (averaging 73 ms in duration), but all were clearly within the aspirated allophonic category for /p/, /t/ and /k/. Participants' shadowed responses to the model's non-lengthened VOTs were 61 ms in Experiment 4a and 69 ms to the lengthened VOTs. The corresponding values were 53 and 57 ms in Experiment 4b, again, all clearly characteristic of aspirated VOTs in English. (To use the examples of Mitterer and Ernestus, the values in Fowler et al. (2003) were characteristic of the VOT in "use pies," not of that in "you spies.")

In short, in our view, theoretical accounts of speech perception that invoke perception of gestures provide a superior account of the present findings and those of Fowler et al. (2003) than do other extant accounts of speech perception. Certainly, we do not contend that other accounts could not be modified especially to "post-dict" our findings. Where gestural theories are concerned, we prefer the direct realist account over the motor theory for two reasons. First, we judge the motor theory's invocation of analysis-by-synthesis to explain extraction of gestural information from acoustic speech signals to be implausible and unnecessary. In favor of the motor theory, however, direct realist theory has not heretofore (e.g., Fowler, 1996) found it necessary to invoke speech motor involvement in speech perception, yet findings of Fadiga et al. (2002) and especially of D'Ausillo et al. (2009) have recently shown motor involvement that implicates perception. Second, we question motor theory's claim that, in respect to perception of gestures (distal, not proximal events), speech perception is special. Although recent evidence from cortical activation may suggest that there are circumstances under which speech and nonspeech stimuli can be processed differently (Whalen et al., 2006), we remain skeptical of the need to invoke modularity in general.

In the present study, although it is clear that all imitators were able to perceive and reproduce aspects of the model's articulation that were intentionally controlled by the model, not every controlled aspect of the articulation was copied faithfully. By way of explanation, we suggest that some forms of gestural organization, while perceived under rapid shadowing, are not easily imitated by naïve talkers (at least in the absence of explicit articulatory training; see, for example, Catford & Pisoni, 1970). In other words, we attribute the small size of the effect to a behavioral conflict in the imitators. Specifically, we suggest that the instruction to shadow speech sets the disposition to imitate into competition with the listener's own practiced speech habits. Even when the shadower fulfills the disposition toward mimesis of positionally unpredictable lateral chromatics as perceived, that listener must overcome the tendency to rely upon one or the other practiced pattern of coordination among gestural constellations for allophones of /l/. In the present experiments, listeners perceived and in a small way reproduced those 'abnormal' events even when doing so required them to combat their own highly practiced speech motor routines as well as any tendency to classify laterals based on syllable-affiliation.

5.3. *Further reflections*

Humans imitate the behavior of others quite generally (e.g., Chartrand & Bargh, 1999; Wilson, 2001), so it should not be surprising that imitation occurs when the behavior in question is the articulation of speech. The human disposition to imitate raises the question, "What supports the human ability to imitate?" The following preliminary answer finds support in the present results.

Infants attempt to imitate from birth. Meltzoff and Moore report that neonates imitate facial gestures even as young as 42 min of age (e.g., 1999). That they do so is remarkable. As Meltzoff and Moore note, young infants can see the tongue of the model when the model produces a tongue protrusion gesture, but they cannot see their own tongues. They can feel their own tongues, but they cannot feel the model's tongue. How do they know which of their own body parts corresponds to the protruding tongue of the model? Meltzoff and Moore suggest that perception yields a *supramodal representation*—a representation that transcends sensory modalities. Infants represent the model's tongue based on optical information and represent their own tongues based on somatosensory information. Because the representations are of distal world properties rather than of proximal sensory patterns, they can equate their own tongues with the tongue of the model.

Remarkably, infants can also match speech across modalities. Presented auditorily with a single vowel (/a/ or /i/) and with two films, one displaying a face mouthing /a/ and one displaying a face mouthing /i/, infants tend to gaze longer at the face mouthing the vowel they hear (Kuhl & Meltzoff, 1982; see also MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). If Meltzoff and Moore's account of infants' imitative ability has generality, it may explain the matching of vowels with faces as well. Essentially, one might argue that infants develop a supramodal representation of distal events that they learn from proximal stimulation. They develop a distal representation of a speaker producing, say, /a/, from optical information they obtain from one of the films; they develop a distal representation of the same vowel from acoustical information. The integrated representation of perceptions of distal speech events allow the infant to perform successful cross-modal matching. Perceiving articulation also allows the infant, eventually, to learn to talk by perceiving model speech.

We ascribe the adult's ability to imitate speech to similar causes. Adults extract articulatory information from the speech they hear, and, if they are disposed to imitate, they imitate its gestures. Gestural imitation produces acoustic similarities between model and imitator.

The present work confirms that human listeners are able to perceive and, in a limited way, reproduce gestures isolated from the spatiotemporal constellations in which they normally appear, even when those gestures may be phonetically aberrant (with respect to the shadowers' own usual positional variants), phonologically aberrant (in terms of syllable position) and semantically aberrant (in terms of being embedded in non-words). We base this claim on the fact that our imitators clearly used the tongue to distinguish the lateral variants in Experiment 3 even though they had to ignore their own system's linguistic constraints and overcome the weight of highly practiced speech motor routines in order to do so. The latter obstacle—conflicting linguistic practice—likely explains why imitative gestural fidelity was far below ceiling. The key point, however, is that even when detailed vocal tract shape must be recovered because the stimuli cannot be mapped easily into the equivalence classes of one's own system, shadowers managed to recover and reproduce at least some of the relevant gestural information. Although listeners may be less sensitive to events that are not normally meaningful for them in their navigation of their environment, and although they may experience some degree of awkwardness in reproducing events that involve unrehearsed motor routines, they are aware at some level that movement of the vocal tract has acoustic consequences in the real world, and are able to recover the source of the sound without previous experience producing it. Acoustic/auditory theorists would of course be able to explain our data satisfactorily by simply saying that it was F2 or F2–F1 that our shadowers perceived, but that they happened to use the tongue in shadowing it even though they had other options. We think it unlikely that listeners did not hear that it was the tongue not the lips producing this contrast, particularly because the model's F1 values would have been lower had he rounded the lips rather than retracted the tongue.

Peer-review of the present work raised a question about the nature of the stimuli. Specifically, the review team was concerned that the exaggerated nature of the stimuli in Experiments 2 and 3, along with the very minimal linguistic contexts provided throughout, may have led the shadowers to focus on whatever differences there were between the stimuli far more than they might when listening to normal speech.

Our response is that, indeed, we intended to force the shadowers to focus on whatever gestures were present more than they would in normal speech. To this end, in all three experiments we set out to make some stimuli which, when compared with normal laterals, were abnormal in terms of syllable position when 'dark'. While we have no reliable empirical measure of syllable affiliation to report (and cannot conceive of how we would make such measures for the approximately unigestural laterals of Experiments 2 and 3), we are confident that our stimuli were all unambiguously V.CV as intended. We base this assertion not merely on our strong intuitions as native speakers of varieties of US English that sport light/dark positional variants of laterals, but also on the fact that we primed our shadowers to perceive the consonants as syllable onsets (see the instructions in Appendix A). In Experiments 2 and 3, our model stimuli were not only positionally uniform as before, but also chromatically enhanced by the model talker. Our aim throughout was to discourage participants from substituting positional variants from their own allophonic inventories if it was possible for them to reproduce the unfamiliar target utterances more directly.

The review team also raised the possibility that the shadowers did not hear the lateral chromatics at all, but simply heard syllable boundaries and substituted their own positionally appropriate bigestural allophone ([l] for V.CV and [ɫ] for VC.V) on the basis of syllable boundaries. Again, our native-speaker intuition is that there were no syllable affiliation cues present on which shadowers could have based such a strategy, and we have presented some articulatory evidence for our position. Specifically, for Experiment 3, we confirmed that the model's laterals were at least as unigestural as intended; bigesturality with tip-lag being one correlate of coda position for laterals that was simply missing here (see Browman & Goldstein, 1995). Furthermore, it seems unlikely that subjects could engage in a two-step process of perceiving syllable affiliation then selecting the appropriate allophone; the nature of the task, rapid shadowing, does not lend itself to potentially top-down post-perceptual judgment (see Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982).

More to the point, substitution simply did not occur; for nearly all participants across all three studies, differences in imitated lateral chromatics (light/dark differences in F2–F1) were much closer than one would expect had participants been substituting their normal lateral allophones, even in Experiment 1 where substitution would have been a more reasonable strategy because the target laterals were strange primarily in position only and then only half the time (see Fig. 3). Furthermore, had there been allophonic substitution, one would have expected some shifting of syllable boundaries to accompany it. To our ears, none of the subjects ever produced a shifted syllable boundary. The lack of substitution should not be surprising; on debriefing, shadowers indicated that they recognized the stimuli as speech as instructed, and yet also recognized the abnormality of the stimuli, thus implying awareness that they were not hearing normal allophones of /l/.

Although all listeners who imitated were sensitive to linguistically irrelevant gestures and reproduced them, not all matched the model as closely as others, and some did not imitate at all. Direct realism posits greater sensitivity to aspects of our environments that matter to us. Our strange /l/ variants in the present set of experiments may have been more familiar to some listeners than others (who may have been exposed to a greater variety of dialects outside the laboratory), or it may simply have been that some of our imitators were more sensitive perceivers of speech or more talented producers of it. We know that some adults are better at second accents, impressions and imitative tasks than others (see, for reviews, Markham, 1997; Zetterholm, 1997). Unfortunately, we lacked a tool that would allow us to pre-screen for proclivity to imitate. We will pursue this question in future work.

Acknowledgements

The present research was supported by National Institutes of Health Grant NIH DC-03782 and NIH Grant DC-02717, both to Haskins Laboratories. The authors thank Gerry Docherty and two anonymous reviewers for very helpful comments as well as Len Katz, Gordon Ramsay and Doug Whalen who critiqued aspects of the work at various stages. For assistance with the instants-based spectrum estimation of Experiment 1, the authors are indebted to B. Yegnanarayana and S. R. Mahadeva Prasanna. Interpretation of the data are our own.

Appendix A. Participant instructions

For each trial in this experiment, you will hear a 2 second warning tone followed by a voice saying various words. Each word will begin with the vowel 'aaahh', and will then switch to a consonant–vowel syllable (for example, 'aaahh–ra'). What you need to do is keep up with the voice. As soon as you hear the vowel on a given trial, immediately begin saying it. When the voice switches to the consonant–vowel syllable, you should switch to the same syllable as soon as possible. Therefore, you are saying the word AS YOU HEAR IT.

Sometimes people race through a word too fast and get short of breath. This being the case, it is important that you take some time before and during the warning noise to take a deep breath in preparation for saying the upcoming word. Also, if you cannot hear the consonants well, please let the experimenters know so that the volume can be adjusted.

There is a total of 288 trials.

References

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America*, *63*(5), 1535–1555.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, *36*, 373–383.

Benus, S., & Gafos, A. (2007). Articulatory characteristics of Hungarian "transparent" vowels. *Journal of Phonetics*, *35*, 271–300.

Borg, G. (1946). Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe: Bestimmung der Differentialgleichung durch die Eigenwerte. *Acta Mathematica*, *78*, 1–96.

Bourhis, R. Y., & Giles, H. (1977) The language of intergroup distinctiveness. In H. Giles (Ed.), Language, ethnicity and intergroup relations (pp. 119–135). London, England: Academic Press.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, *49*, 155–180.

Browman, C., & Goldstein, L. (1995). Gestural syllable position effects in American English. In F. Bell-Berti, & L. J. Raphael (Eds.), *Producing speech: Contemporary issues* (pp. 19–33). New York: AIP Press.

Catford, J., & Pisoni, D. (1970). Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal*, *54*(7), 477–481.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Chiba, T., & Kajiyama (1941). *The vowel: Its nature and structure*. Tokyo: Tokyo-Kaiseikan Publishing Company, Ltd.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141–177.

D'Ausillo, A., Pulvermueller, F., Saimas, P., Bufalari, I., Begllomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385.

Delattre, P. (1971). Consonant gemination in four languages: An acoustic, perceptual, and radiographic study, part I. *International Review of Applied Linguistics in Language Teaching*, *9*(1), 31–52.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American /r/s by X-ray motion picture. *Linguistics*, *44*, 29–68.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179.

Faber, A. (1989). On the nature of proto-semitic *l. *Journal of the American Oriental Society*, *109*(1), 33–36.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*, 399–402.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1730–1741.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*, 396–413.

Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language*, *58*, 815–836.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *13*, 361–377.

Galef, B. G., Jr. (1988). Imitation in animals: History, definition, and interpretation of data from the psychological laboratory. In T. R. Zentall, & Galef, Jr. (Eds.), *Social learning: Psychological and biological perspectives* (pp. 3–28). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gaskell, G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 144–158.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI: Constraints on phonetic interpretation* (pp. 222–236). Cambridge, England: Cambridge University Press.

Gick, B., A. Kang, M., & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*, 357–371.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.

Giles, S. B., & Moll, K. L. (1975). Cineflourographic study of selected allophones of English /l/. *Phonetica*, *31*, 206–227.

Goldinger, S. (1996). Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

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

Gopinath, B., & Sondhi, M. M. (1970). Determination of the shape of the human vocal tract from acoustical measurements. *Bell Systems Technical Journal*, *49*, 195–1214.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, *105*(5), 2854–2865.

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Heinz, J. M. (1967). Perturbation functions for the determination of vocal-tract area functions from vocal-tract eigenvalues. *STL-QPSR*, *8*(1), 001–014.

Hogden, J., Lofqvist, A., Gracco, V., Zlokarnik, I., Rubin, P., & Saltzman, E. (1996). Accurate recovery of articulator positions from acoustics: New conclusions based on human data. *Journal of the Acoustical Society of America*, *100*(3), 1819–1834.

Hogden, J., Valdez, P, Katagiri, S., & McDermott, E. (2003). Blind inversion of multidimensional functions for speech enhancement. In *EUROSPEECH 2003* (pp. 1409–1412). [On-line]. Available: 〈www.isca-speech.org/archive/eurospeech_2003〉.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, *108*, 710–722.

Honorof, D. N. (1999). Articulatory gestures and Spanish nasal assimilation. *Dissertation Abstracts International*, *60* (12A), 4403 (University Microfilms No. 99-54317).

Honorof, D. N., & Browman, C. P. (1995). The center or edge: How are consonant clusters organized with respect to the vowel? In K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 3 (pp. 552–555). Stockholm, Sweden: KTH and Stockholm University.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, *8*, 164–181.

Johnson, K. (1997a). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics*, *50*, 101–113.

Johnson, K. (1997b). *Acoustic & auditory phonetics*. Malden, MA: Blackwell.

Jones, D. (1909/1962). *The pronunciation of English (Third Impression; Fourth American Edition)*. Cambridge, England: Cambridge University Press.

Kac, M. (1966). Can one hear the shape of a drum? *American Mathematical Monthly*, *73*(4), 1–23.

Kozhevnikov, V. A., & Chistovich, L. A. (1965). *Speech: Articulation and perception* (U.S. Department of Commerce, Trans.). Washington, D.C.: Joint Publications Research Service (No. 30), p. 543.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.

Kuhl, P., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*, 2425–2438.

Labov, W. (1963). The social motivation of a sound change. *Word*, *XIX*, 273–309.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. *Cognition*, *21*, 1–36.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, *7*, 147–161.

Lunn, J., Wrench, A. A. & Mackenzie Beck, J. (1998). Acoustic analysis of /l/ in Glossectomees. *SST Student Day Poster presented at the 5th International Conference on Spoken Language Processing, (ICSLP 98)*, Sydney, Australia. Available: 〈sls.qmuc.ac.uk/pubs/lun981.pdf〉.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219*, 1347–1349.

McHugo, G., Lanzetta, J., Sullivan, D., Masters, R., & Englis, B. (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality and Social Psychology*, *49*, 1513–1529.

Markham, D. (1997). Phonetic imitation, accent and the learner. *Travaux de l'institut de linguistique de Lund, 33*.

Meltzoff, A., & Moore, M. (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel, & G. Butterworth (Eds.), *Imitation in infancy* (pp. 9–35). Cambridge, England: Cambridge University Press.

Mermelstein, P. (1967). Determination of the vocal tract shape from measured formant frequencies. *Journal of the Acoustical Society of America*, *41*, 1283–1294.

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

Nagell, K., Olguin, K., & Tomasello, M. (1993). Processes of social learning in tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Journal of Comparative Psychology*, *107*, 174–186.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *Journal of the Acoustical Society of America*, *101*(2), 1078–1089.

Palmeri, T., Goldinger, S., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *19*, 309–328.

Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In W. Labov (Ed.), *Locating language in time and space* (pp. 143–178). New York: Academic Press.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, *92*(6), 3078–3096.

Pisoni, D. B., & Lazarus, J. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, *55*, 328–333.

Porter, R., & Castellanos, F. X. (1980). Speech production measures of speech perception: Rapid shadowing of VCV syllables. *Journal of the Acoustical Society of America*, *67*, 1349–1356.

Porter, R., & Lubker, J. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage. *Journal of Speech and Hearing Research*, *23*, 593–602.

Rickford, J. R., & McNair-Knox, F. (1994). Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 235–276). Oxford, England: Oxford University Press.

Riordan, C. J. (1977). Control of vocal-tract length in speech. *Journal of the Acoustical Society of America*, *29*, 1462–1464.

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

Schroeder, M. R. (1967). Determination of the geometry of the human vocal tract by acoustic measurements. *Journal of the Acoustical Society of America*, *41*(4), 1002–1010.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*, 489–573.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation of shadowed words. *Perception and Psychophysics*, *66*, 422–429.

Sondhi, M. M. (1979). Estimation of vocal-tract areas: The need for acoustical measurements. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *27*(3), 268–273.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Tabachnick, B. G., & Fidell, L. S. (1989). *Chapter 11: Discriminant function analysis. Using multivariate statistics (second ed.)*. New York: Harper & Row, Publishers pp. 505–596.

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

Tomasello, M. (1996). Do apes ape?. In C. M. Heyes & B. G. Galef, Jr. (Eds.), *Social learning in animals: The roots of culture* (pp. 319–346). San Diego, CA: Academic Press.

Wells, J. C. (1982). *Accents of English (3 volumes)*. Cambridge: Cambridge University Press.

Westbury, J. R. (1994). *X-ray microbeam speech production database user's handbook*. Madison WI: X-ray Microbeam Facility.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, *25*, 501–528.

Whalen, D. H., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., & Lai, S., et al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America*, *119*, 575–581.

Whiten, A., & Custance, D. M. (1996). Studies of imitation in chimpanzees and children. In C. M. Heyes, & B. G. Galef, Jr. (Eds.), *Social learning in animals: The roots of culture* (pp. 291–318). San Diego, CA: Academic Press.

Wilson, M. (2001). Perceiving imitatible stimuli: Consequences of isomorphism between input and output. *Psychological Bulletin*, *127*(4), 543–553.

Yegnanarayana, B., & Veldhuis, R. N. J. (1998). Extraction of vocal-tract system characteristics from speech signals. *IEEE Transactions on Speech and Audio Processing*, *6*(4), 313–327.

Yehia, H. C. (1997). *A study on the speech acoustic-to-articulatory mapping using morphological constraints*. Unpublished Doctoral Dissertation. Nagoya, Japan: Graduate School of Engineering of Nagoya University.

Zentall, T. & Akins, C. (2001). Imitation in animals: Evidence, function and mechanisms. In R. G. Cook (Ed.), *Avian visual cognition*. [On-line]. Available: 〈www.pigeon.psy.tufts.edu/avc/zentall/〉.

Zetterholm, E. (1997). *Impersonation: A phonetic case study of the imitation of a voice*. Working Papers 46, 269–287. Lund, Sweden: Department of Linguistics, Lund University.