EFFECTS OF IMITATION ON THE ARTICULATION OF CHALLENGING SPEECH TARGETS

Douglas N. Honorof
Yale University & Haskins Laboratories, New Haven, Connecticut, USA

ABSTRACT

Magnetometrically transduced articulator position data were collected for pre-vocalic American English /r/ and /l/ as spoken by four adult Japanese learners of English under native-speaker imitation and non-imitation conditions. Japanese has only one liquid phoneme, therefore imitation of these contrasting and reportedly difficult English sounds was expected to decrease token-to-token variability and to increase overall distinctiveness of sets of articulator positions for the two members of the contrast. Stepwise discriminant analyses and variability testing provide support for two claims. 1) The visual modality may be relied upon during imitation when the information it provides is especially salient. 2) Imitation of a native-speaker model facilitates articulatory control over challenging L2 speech contrasts in inverse relation to previous mastery of control for the individual targets.

1. INTRODUCTION

Imitation is a commonly used technique in articulation therapy and in the teaching of L2 pronunciation, articulatory phonetics and speech for dramatic purposes. However, the effect of the act of mimicry on articulation itself is not yet well understood. The present multi-speaker magnetometer and acoustics study directly investigates the immediate effects of imitation on the articulation of non-native linguistic targets.

The present work proceeds from the assumption that the ability of adult second accent learners to reproduce the information in a modeled speech signal should be enhanced while that signal remains fresh in short-term memory. One way of quantifying degree of success in reproduction of a speech signal would be to examine the vocal tract configurations themselves. Because of anatomical differences between speakers, such cross-speaker comparison of productions is not a straightforward matter. However, Naito has found token-to-token variability of articulator position to be one dimension along which speakers differ most in speaking L1 and L2 [1]. The experimental results reported here allow us to address control over L2 targets for liquids under imitation and non-imitation conditions in terms of both variability and overall distinctiveness of midline oral articulator positions.

2. METHODS

2.1. Design and Stimuli

The present study forms a subset of a mini-longitudinal study incorporating a training component and a larger set of stimuli. Due to limitations of space, only the pre-training results for one vowel context are reported here.

The present design varies liquids in absolute utterance-initial position. These sounds were chosen because differences between /r/ and /l/ are sub-phonemic in Japanese, because the Japanese liquid phoneme has been described as often being articulated very differently from either English liquid, and because Japanese learners of English have been reported to have difficulty acquiring this contrast, especially in perception [e.g., 2, 3, but see 4, 5].

English stimuli—*RAY dough* and *LAY dough*—were presented in conventional English orthography. Subjects were instructed to emphasize words presented in uppercase lettering. A phonologically analogous Japanese utterance, *reido desu* ('It is zero degrees.') was chosen for comparison. It is given here in a Romanized script, but was presented to the subjects in *kana*. All stimuli were iterated 15 times and randomized along with other English utterances. Stimuli were presented in blocks by condition, with the Japanese condition being presented last.

2.2. Data Collection Technique and Procedure

Subjects were situated in an articulometer [6] facing the author, a native speaker of American English, who was seated immediately in front of the subjects. The Japanese speakers were instructed to observe the American's speech carefully by eye and ear as he read each utterance aloud, and then to produce imitations of the American's utterances while simultaneously reading the utterances from print. It was hoped that providing information from multiple sources would afford to the speakers a relatively high possibility of successful imitation. Before and after the imitation condition, subjects read aloud another 15 randomized iterations of each of these utterances without the assistance of the native-speaker model.

2.3. Subjects

Four native speakers of Japanese participated in the experiment. Two had spent most of the previous year in the United States. One of them was an ESL student during this time. During the period immediately preceding the experiment, three had spent over one month in the United States where they were studying English in an intensive ESL program. Prior to their arrival in the United States, some had traveled briefly to English speaking destinations. All had studied English primarily from textbooks from early adolescence, though, subjectively judged, none spoke English very fluently. Three of the subjects were male, one female. They ranged in age from 28 to 41. None reported any speech, hearing or language impairment. All subjects were familiar with the English speech of the L2 model from roughly equivalent amounts of previous contact.

2.4. Measurement Procedure

Because subjects were expected to exhibit a lack of consistent control over /r/ and /l/ and to adopt highly individual strategies for producing the sounds, no attempt was made to identify a

stable articulatory index for any of the gestures employed in the production of English liquids. Rather, onset of voicing was used to index attainment of consonant target. Articulator positions at voicing onset were measured for eight transducer coils in two dimensions, providing 16 measurement points in all. These coils were affixed with adhesive to the midline surfaces of the following flesh points: under tongue tip (*UTT*), post-tongue tip (*TT*), tongue blade (*TBL*), tongue center (*TC*), tongue dorsum (*TD*), the vermilion borders of the upper and lower lips (*UL* and *LL*, respectively), and to the gum line between the lower incisors to index mandibular movement (*MB*). In general, onset of voicing lined up very well with peaks in tongue tip vertical, and, to a lesser extent, horizontal, movement curves. For one subject, onsets of voicing were identified with reference to spectrographic landmarks due to the presence of machine-related noise in the audio signals that made it difficult to read voicing onset directly from the waveforms.

3. RESULTS

3.1. Distinctiveness: Single-speaker Multivariate Analyses

If we were to plot the measurement points at onset of voicing during target consonant articulations on a grid laid over a midsagittal face diagram, our familiarity with such diagrams and our skill at visual pattern recognition might allow us to "eyeball" any differences in tongue shape from consonant to consonant. These overall differences, however, might not be apparent had we plotted only a single dimension of a single articulator. For example, even without considering tongue shape, one can imagine lip aperture and protrusion along with jaw height as together distinguishing /r/ and /l/ from each other despite token-to-token differences in the contributions of the individual articulators (lips and jaw) along a single dimension. Here the assumption is that, while it may be possible to control the vertical movement of the upper lip, for example, relatively independently of the movement of the jaw and lower lip, in speech it is often the case that slightly more global combinations of articulator contributions adhere into gestural units [7, 8]. Fortunately, with discriminant analysis, it is possible to assess the extent to which a number of parameters work together to contribute to the overall distinctiveness of the data, but in a way that does not make *a priori* assumptions about which articulators are involved in the target consonant gestures apart from the assumptions underlying the selection of flesh points for the transducer coils.

Conceptually, for each subject, the present stepwise discriminant analysis reduces 16 input dimensions (one for each articulator/dimension) to a single derived dimension in which the two levels of the grouping variable, here imitated /r/ and imitated /l/, are maximally distinct. (For a more detailed description of discriminant analysis, see [9].) The imitation condition was chosen as the grouping factor because imitation was predicted to produce the most distinct, and thus the most easily classified, articulations. For all four subjects, the discriminant analysis confirmed that imitated /r/ and /l/ were significantly different. The distances between the solution of the discriminant function for each token and group means for imitated /r/ and for imitated /l/ group are also calculated. On this basis a token is classified as a member of one of the two groups independent of its actual consonant target.

With a single exception, jackknifed classification of the articulations of three of the four speakers always patterned correctly along phonological/orthographic lines under all three English conditions. The exception was, for MT, one out of 15 /l/s under imitation was classified as /r/. For Subject YH, however, one of the 14 pre-imitation /l/s was classified as an /r/ and seven out of the 15 pre-imitation /r/s were classified as /l/.

The jackknifed classification of the Japanese articulations tended to group the L1 liquid with L2 /l/. Of the three subjects for whom L1 data are available, only MT produced any Japanese liquids that were classified as /r/ (4 out of 15). This pattern of classifications alone cannot indicate whether a speaker's English /r/ or /l/ is identical to that speaker's Japanese liquid in the absence of information about the actual distribution of the solutions to the discriminant functions for each token. These solutions, coded by condition, are plotted in the derived space in Figure 1, where the number of iterations varies from 13 to 15 for each of the seven conditions with the exception of the plot for YH, whose data set lacks post-mimicry and Japanese conditions. On examining this figure, the reader may be struck by the degree to which the overwhelming majority of MT's Japanese liquids actually lie between /r/ and /l/ in all English conditions despite the forced classification into one or the other group.

The mean coordinates for partial *F* values are provided by the discriminant analysis for each input variable. These *F* values allow us to rank the input variables according to the magnitude of their contribution to the derivation of the discriminant function, which allows to us interpret the results in terms of conventional assumptions about the articulatory organization of speech. Table 1 lists the coefficients (standardized by pooled within-group variances) for the input variables with the six highest *F*s-to-remove at the final step of the analysis for each subject. The higher the absolute value of the coefficient for a given articulator/dimension, the greater its contribution to the derivation of the canonical variable, and thus to the overall definition of the derived space.

3.2. Articulatory Control: Univariate Variability Testing

In variability testing we are uninterested in testing the null hypothesis of no difference between treatments. Rather, we want to know whether the vertical movement of the tongue tip at the beginning of the word *ray*, for example, is achieving the same idealized target on each iteration *more consistently* under one condition (imitation, by hypothesis) than under the others. Therefore, one-way equality of variance tests were run on measurements from each dimension (x and y) of each of the eight coils for each English liquid (/r/ and /l/). These tests examine the absolute deviation of each measurement point from its group mean, then compare the overall variability around group means for pre-imitation versus imitation, for imitation versus post-imitation, or for pre-imitation versus post-imitation. Thus there were three comparisons for each of 16 independent measurement points for each of two English utterances (n=15 except as noted above). With multiple tests, there is always the risk that one or more will be significant by chance. Therefore, the confidence level for

Subject	Input Variable	Input Variable	Input Variable	Input Variable	Input Variable	Input Variable
YH	tty	utty	mby	tdx	tcy	tbly
	2.9	-3.5	-2.2	1.3	-2.2	2.1
HM	ulx	llx	mbx	tdy	uttx	uly
	-2.1	3.6	-2.5	1.8	1.2	-0.6
MM	lly	mby	tdy	ttx	tbly	tdx
	1.6	-1.3	1.9	2.3	-2.0	1.4
MT	uly	mby	utty	tbly	tdx	llx
	0.8	-1.9	1.2	1.9	2.0	-1.1

Table 1. Results: Input variables contributing most reliably to the canonical variable produced by /r/-/l/ discriminant functions for each speaker, together with standardized coefficients. Input variables are ranked left-to-right by *F*-to-remove at final step. Negative numbers indicate low or back.

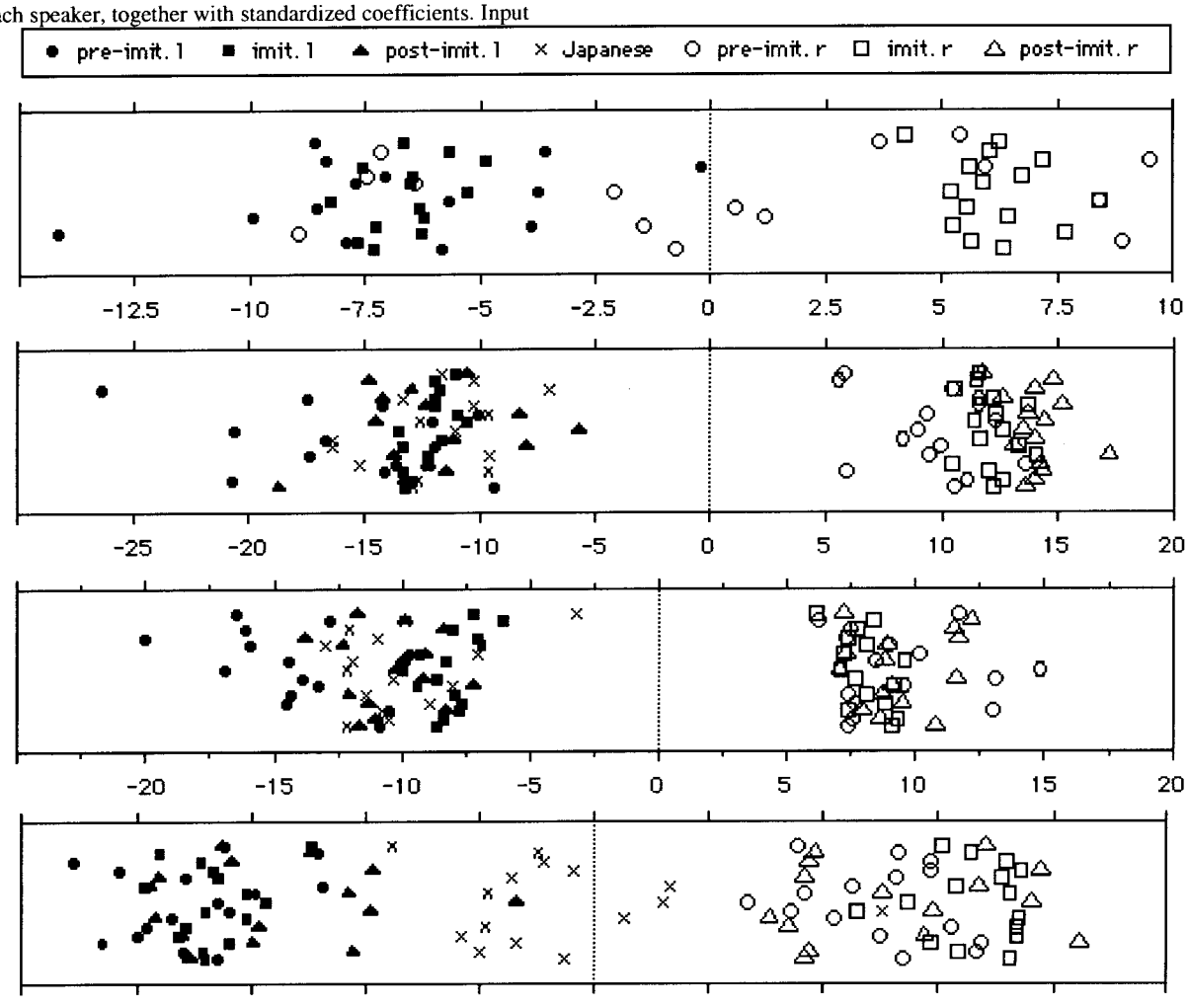

Figure 1. Univariate scattergrams of solutions to a 16-term linear discriminant function in derived space for four speakers with *imitation* as the grouping factor. With only two levels to the grouping factor (/r/ and /l/), all the dispersion in the data is accounted for by a single discriminant function. Solutions are coded by condition.

these tests was adjusted to $p < .0166$ (i.e., $p < .05/3$). A Levene's p value reaching this confidence level indicates a significantly lower degree of variability around the cell mean for the level of the grouping variable with the smaller of the two standard deviations. Table 2 lists Levene's p values by target word and by subject. Only significant results are included.

4. DISCUSSION & CONCLUSIONS

The discriminant analysis shows that all four subjects attempted to distinguish /r/ from /l/ before during and after imitation with varying degrees of success. The fact that these speakers' /r/ and /l/ articulations were generally distinguished in the analyses does not imply that their articulations were native-like, however.

There is an overwhelming tendency for the Japanese liquids produced by these speakers to be classified as /l/ in the discriminant analyses. Given that there were also so few differences in token-to-token variability for /l/, perhaps /l/ is simply the easier liquid for these speakers to produce. However, given that all but one of MT's Japanese liquids lie between /r/ and all but one /l/ in the derived space, we should not conclude that there is a simple substitution of the L1 liquid target for an English /l/, at least not for this subject. Although at first glance this subject's /l/ values appear to spread over a greater range than for the other subjects, the scale is much smaller, and the range of values actually smaller. It would appear that this subject was the one with the greatest distinctiveness between L1 and L2 targets in all conditions (though not the greatest distance between L2 means in derived space), and is the only subject in the present study for whom the equality of variance testing of individual measures showed a repeating pattern of counter-facilitation during imitation and never the opposite pattern. Note that this subject's articulations stabilize again after imitation.

The Levene's tests provide a less clear pattern for the other three subjects, and are inconclusive with regards to the original hypothesis that imitation facilitates control. However, the plots from the discriminant analyses are suggestive of a substantial

gain in overall control over L2 targets during imitation for subject YH, and to a lesser extent for the other three subjects as well. For HM, the imitation and post-imitation /r/ become more distinct (i.e., are shifted rightward in derived space). The pre-imitation/post-imitation comparisons are difficult to interpret, however, because there is an inherent confound in this type of design between learning, increased familiarity with the task or stimuli over time, and increased skill at speaking past the transducer coils. Had the these comparisons shown greater stability before training overall, the pre-imitation/post-imitation comparisons would have been more informative.

Looking at just the first few articulator/dimensions, i.e., those that contribute most reliably to the discriminant functions (Table 1), it would appear that /r/ and /l/ are being distinguished most strongly on the basis of gestures involving retroflexion (especially for YH: TTY is high when UTTY is retracted), some combination of lip protrusion and jaw lowering (MM and MT) or some other less straightforwardly interpreted combination of lip and jaw activity. One possible conclusion here is that the lips, tongue tip, under tongue tip and jaw are especially visible for these sounds, allowing information available through the visual modality to used by all four speakers, perhaps along with auditory and articulatory/proprioceptive information [See 10, 11, 12].

Although there are respects in which all four subjects behaved similarly, the lack of uniformity in one respect may actually hint at a coherent story. If we consider Subject MT to have begun with the greatest control in terms of distinctiveness between L1 and L2 liquids, we may speculate that the variability seen for so many individual measures for imitated /r/s follows from this speaker's superior control of that sound before imitation. Perhaps this speaker was not able to identify subtle differences between his /r/s and the model's, and thus arrived at no consistent strategy for changing. It is also possible that this speaker was simply attempting with difficulty to shift from one type of /r/ which he already controlled to the one being modeled [13]. At the other extreme, YH seems to have started out with the greatest confusion, especially as seen in the number of misclassified pre-imitation /r/s, but also seems to have gained considerable consistency from token to token during imitation. (See Figure 1). Given these findings, it may be that imitation facilitates articulatory control over challenging L2 targets in inverse relation to previous target mastery.

In general the present results provide support for the notion that imitation can have an effect on production in the immediate term, but the specific effect may depend on input proficiency. No claim is made here regarding carryover effects into improvement in production on demand. (See [14] for a discussion of these issues.)

ACKNOWLEDGMENTS

The author thanks Walter Naito for his assistance during the early stages of the present project and Alice Faber for her coaching on discriminant analysis. Neither is responsible for any shortcomings of the final product. While preparing the present paper, the author received support from NSF Grant SBR-9514730 and NIH Grant HD-01994 to Haskins Laboratories.

REFERENCES

[1] Naito, W. R. 1995. English /r/ and /l/ production in American and bilingual Japanese subjects. *JASA,* 98 (5), 2892 (A).

[2] Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia,* 9, 317-323.

[3] Sheldon, A. and Strange, W. 1982. The acquisition of /r/ and l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics,* 3, 243-261.

[4] Best, C. T. and Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics,* 2, 305-550.

[5] Bradlow, A. R., Pisoni, D. B., R. Akahane-Yamada and Y. Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *JASA,* 101 (4), 2299-2310.

[6] Perkell J. et al. 1992. Electromagnetic Midsagittal Articulometer (EMMA) systems for transducing speech articulatory movements. *JASA,* 92, 3078-3096.

[7] Browman, C. P. and Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica,* 49, 155-180.

[8] Mattingly, I. G. 1990. The global character of phonetic gestures. *Journal of Phonetics,* 18, 445-452.

[9] Dixon, W. J. (ed.) 1992. *7M: Stepwise Discriminant Analysis.* BMDP Statistical Software Manual Volume 1, 363-385.

[10] McGurk, H. and MacDonald, J. 1976. Hearing lips and seeing voices. *Nature,* 264, 746-748.

[11] Catford, J. C. and Pisoni, D. 1970. Auditory vs. articulatory training in exotic sounds. *Modern Language Journal,* 54, 447-481.

[12] Markham, D. 1997. Phonetic Imitation, Accent, and the Learner. *Travaux de l'Institute de Linguistique de Lund,* 33. Lund: Sweden.

[13] Delattre, P. C. and Freeman, D. C. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics,* 44, 29-68.& Freeman

[14] Landahl, K. L. and Ziolkowski, M. S. 1995, ms. Discovering phonetic units: Is a picture worth a thousand words? To appear in A. Dainora et al. (eds.), *Papers from the 31st Regional Meeting of the Chicago Linguistic Society. Vol. 1: The Main Session.* Chicago: CLS.

target	*subject*	*artic./ dimen.*	*comparison*	*larger SD*	*Levene's* p
	YH	lly	pre / imit	imit	$p < .0158$
	HM	utty	pre / post	pre	$p < .0002$
		ttx	imit / post	imit	$p < .0157$
		tbly	imit / post	imit	$p < .0111$
		llx	pre / imit	pre	$p < .0009$
		"	pre / post	pre	$p < .0062$
	MM	uttx	imit / post	imit	$p < .0010$
		"	pre / post	pre	$p < .0026$
		utty	pre / post	post	$p < .0122$
		mbx	pre / post	pre	$p < .0152$
R		mby	pre / imit	pre	$p < .0005$
A		"	pre / post	pre	$p < .0011$
Y		ulx	pre / post	pre	$p < .0146$
	MT	uttx	pre / imit	imit	$p < .0009$
		"	pre / post	post	$p < .0099$
		ttx	pre / imit	imit	$p < .0063$
		tbly	pre / imit	imit	$p < .0020$
		"	imit / post	imit	$p < .0056$
		tcx	pre / imit	imit	$p < .0041$
		"	imit / post	imit	$p < .0133$
		tdx	pre / imit	imit	$p < .0014$
		"	imit / post	imit	$p < .0146$
		tdy	pre / imit	imit	$p < .0009$
		"	imit / post	imit	$p < .0130$
L	*YH*	tdx	pre / imit	pre	$p < .0049$
A	*HM*	ulx	pre / post	pre	$p < .0022$
Y	*MM*	tbly	pre / imit	imit	$p < .0101$

Table 2. Cells with greater variability where Levene's p values reach an adjusted confidence level of $p < .01666$.