

From J. Slifka, S. Manuel, & M. Matthies (Eds.) *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*. (2004).

MAPPING THE TASK FOR THE SECOND LANGUAGE LEARNER: THE CASE OF JAPANESE ACQUISITION OF /R/ AND /L/

Andrew J. Lotto¹, Momoko Sato² & Randy L. Diehl²

¹Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, USA

²Department of Psychology, University of Texas, Austin, TX USA

lottoa@boystown.org

ABSTRACT

The acquisition of a foreign phonetic contrast requires the second language (L2) learner to attend to those acoustic dimensions that are informative for the distinction and to manipulate values along those dimensions during production. The discovery of informative dimensions in L2 can be complicated by the contrasts present in the native (L1) language. A well-known example is the difficulty that native Japanese speakers have perceiving or producing the English /l/-/r/ distinction. Here, we attempt to systematically describe this L2 learning task by obtaining distributions of acoustic measures (formant frequencies and durations) from native English productions of word-initial /l/ and /r/. These distributions include inter-speaker (gender), intra-speaker, and phonetic (vowel environment) variance. These results reveal that F3-onset frequency provides almost complete discrimination between the distributions. Distributions of native-Japanese productions of /l/ and /r/ were also collected. The Japanese distributions can be partially separated on F2 and F3 onset frequency. All of these distributions are compared to a distribution of native productions of the Japanese rhotic flap. The flap distribution overlaps the native /r/ and /l/ distributions in F2xF3 space. Further measures of native productions reveal that the flap is contrasted with Japanese /w/ mainly by the onset frequency of F2. Thus, the Japanese productions of /l/ and /r/ appear to be influenced by both the informative variance in L2 distributions (F3) and by the informative variance in distributions of similar L1 categories (F2).

INTRODUCTION

The goal of the project reported here is to describe the task of acquiring a second language (L2) phonetic contrast within the framework of general perceptual categorization. Our theoretical assumption is that phonetic classes are essentially perceptual categories defined in a multi-dimensional space. The task of the language user, then, is to parse this space into functional equivalence classes specific to the to-be-learned phonetic system. Within this framework, an estimate of the distributions of phonetic classes within the space allows one to derive an optimal decision strategy (or “ideal observer”) and compare it to actual performance. That is, one can describe the phonetic-category learning task at the level of computational theory in the hierarchy proposed by Marr (1982). Here, we present results from an initial attempt to apply this framework to the well-known problem of Japanese listeners acquiring the English /r/-/l/ distinction.

The Task for The Language Learner

A useful way to approach perceptual categorization problems is to conceptualize a multi-dimensional space in which a particular stimulus is plotted as a point, which represents its

values on the dimensions. The classic formant plots, such as vowels plotted as F1 and F2 frequency values, is such a space. Within this space, a phonetic class is represented as a distribution of points (exemplars). The spread of this distribution is a representation of the variability of productions. When multiple phonetic classes are represented, the distributions are situated in separate areas of the space but there will generally be overlap. If one takes into account the transformations of the auditory system then one can move from an acoustic space to an auditory space. Most of these transformations can be represented as a “warping” of the space. That is, one dimension may be stretched or squeezed depending on its representation within the auditory system. For example, a dimension can be plotted on a psychophysical scale, which results in a logarithmic distortion of the physical space.

In order to identify the phonetic category of a sound, the listener can parse the space by placing a boundary or decision criterion within the space. If the distributions of a phonetic contrast are non-overlapping on a single dimension then the listener can perform perfectly on a categorization task by placing the boundary on that single dimension (presuming that the distributions are separate in auditory space once internal noise is taken into account). This is an example of an invariant cue (Stevens & Blumstein, 1981). Most phonetic contrasts cannot be distinguished perfectly on a single dimension. Even if one could rely on a single dimension, it may behoove the listener to use information from multiple cues (dimensions) because any one cue could be masked or be unreliable for a particular speaker. When making an identification using multiple dimensions, the user may “weight” the information from each dimension differentially. That is, the final decision may be more readily influenced by the stimulus’ value on some dimensions and less on others. In some categorization models this weighting is conceptualized as “selective attention” (e.g., Nosofsky, 1986) and can be represented as a shrinking or expansion of the space on each dimension. We will avoid the theoretically-loaded term attention and, instead, refer to the correlation between dimension values and categorization responses as dimension “weighting”. An optimal weighting strategy for a perceiver is a function of the task to be performed, the stimulus set, the noise inherent in the representations of these stimuli, and the variables that one means to optimize (e.g., accuracy, efficiency, robustness). In the case of phonetic categorization, we may presume that the weighting of a dimension should be related to its reliability at distinguishing the members of a phonetic contrast,

So, in order to perform optimally in a phonetic identification task, a language learner must give the greatest weight to those dimensions that are most informative for the contrast in question and this weighting strategy is presumed to develop as one gains experience with the phonetic distributions of the language. One problem facing the L2 learner is that the weighting scheme for their native language (L1) may be inappropriate for the L2. If so, then they must be able to shift weighting strategies given the new language input. This flexibility may be easier to describe than to accomplish.

Japanese Speakers and the English //-/r/ Contrast

The difficulty that Japanese speakers have perceiving and producing the English //-/r/ contrast is well-documented (Goto, 1971; Miyawaki *et al.*, 1975). Most theoretical accounts point to the Japanese flap as culpable for these problems. The flap is usually described as perceptually either midway between English // and /r/ or as slightly more similar to //. Thus, it may interfere with attempts to acquire the L2 categories. In order to develop a picture of the category-

formation task facing the Japanese L2 learner, we collected native productions of the English liquids and the Japanese flap as well as Japanese productions of the English contrast. We attempted to provide a rough estimate of the phonetic distributions to which a listener is typically exposed. In particular, we included variance arising from intra-speaker variability (repeated exemplars by same talker), inter-speaker differences (multiple speakers including both genders), and phonetic environment influences (three vowel environments).

METHOD

Six native American English and six native Japanese speakers (gender equally represented in both groups) participated in the study. The members of the Japanese group were all graduate students who had been in the United States for a duration ranging from 9 months to 5 years. English speakers produced six English two-syllable words that varied in initial consonant (/l/ vs. /r/) and subsequent vowel (/i/, /a/, or /u/). The words were “reading”, “leading”, “rocking”, “locking”, “rooting” and “looting”. Two syllable words were chosen because they better matched the Japanese word list. The point vowels were used because they are common to both languages and they provide a wide range of phonetic environment variance. Japanese speakers produced a set of 3 Japanese words beginning with the Japanese flap in the three vowel environments. They also produced each of the words in the English list. All words were produced three times and the order of words (and order of L1 and L2 productions for Japanese speakers) was randomized. Acoustic analyses were then performed on all tokens. Word-onset and mid-vowel frequency values for the first three formants were measured as well as initial consonant and overall syllable duration.

RESULTS

English Productions of /l-/r/

Figure 1a displays a scatter plot of word-onset F2 and F3 values from the English productions of /l/ and /r/. It is clear from this plot that these phonetic distributions can be distinguished quite well by F3 onset frequency, which has long been considered the primary cue for the contrast (O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957). However, to optimize accuracy one would need to consider F2 onset frequency as well. To quantify these observations, all acoustic measures (frequency in mel, duration in ms) were entered as predictor variables of a point-biserial multiple-regression model with phonemic class (/l/ or /r/) as the dependent variable. Three variables were retained in the final model ($p < .05$ cut-off). F3 onset frequency had a standardized beta weight of 0.938; F2 onset frequency had a weight of 0.277 and mid-vowel F1 frequency had a weight of 0.152. This last variable may reflect differences in F1 transitions between /l/ and /r/ (Dalston, 1975; O'Connor *et al.*, 1957). Clearly, the F2 and F3 beta weights are indicative of the importance of F3 onset for the contrast and the ancillary importance of F2.

One may consider these beta weights as a prescription for the optimal weighting pattern for a listener attempting to categorize /l/ and /r/. That is, listeners (including L2 learners) should heavily weight the onset frequency of F3 and more moderately weight F2 onset frequency.

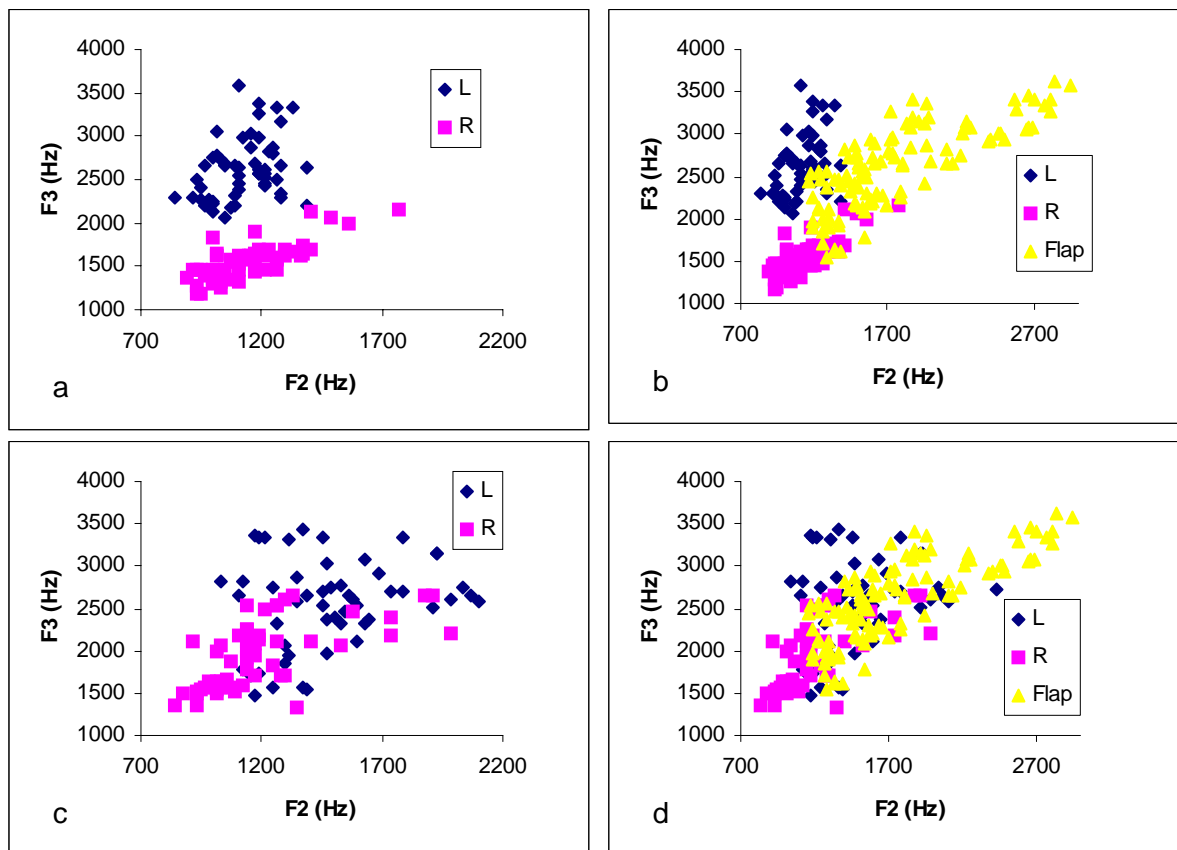


Figure 1. Scatter plots of F2 and F3 onset frequencies obtained from productions of English /l/ and /r/ and the Japanese flap. a) Native English productions of /l/ and /r/; b) English productions of /l/ and /r/ with distribution of native Japanese productions of the flap; c) Native Japanese productions of /l/ and /r/; d) Native productions of flap with L2 productions of /l/ and /r/.

Japanese Productions of Flap

Figure 1b shows the F2 and F3 onset frequencies for native productions of the Japanese flap (yellow triangles) along with the native /l/-/r/ distributions from Figure 1a. Note that the majority of flap exemplars have F3 values that are in the range of English /l/. This is consistent with previous reports that the flap is perceptually more similar to /l/ than to /r/ (Takagi, 1993). There are three salient aspects of the comparisons in Figure 1b that may relate to the difficulty that Japanese speakers have acquiring the English contrast. First, the flap distribution partly overlaps the optimal boundary between /l/ and /r/. Exemplars of both English liquids fall within the flap distribution in F2 x F3 space. That is, Japanese speakers would categorize some exemplars from these two distributions as members of a single category. This would lower the distinctiveness of exemplars near the /l/-/r/ boundary. But the category boundary is just the area of space where exemplars should be more distinctive for proper categorization (acquired distinctiveness, Lawrence, 1949). This reasoning is similar to the category assimilation proposal of Flege (1995) or the perceptual assimilation model (PAM) of Best (1994). However, PAM is based on phonological assimilation and is encumbered with a notion of similarity at the

gestural level. In our approach (and Flege's), one would predict difficulties in L2 category formation based on the overlap of acoustic (auditory) distributions.

A second clue to the L2 problem is that the flap distribution does not extend to lower values of F2. In a follow-up pilot study, we collected Japanese productions of native /w/. The distributions of the flap and /w/ were segregated mainly on F2. That is, the low F2 region of the F2 x F3 space is occupied by the /w/ distribution. It would follow that the optimal weighting strategy for native Japanese listeners would be to weight F2 onset heavily. Perhaps this L1 weighting strategy interferes with the ability to acquire an effective L2 weighting strategy with F3 more heavily weighted (see Iverson *et al.*, 2003; Yamada & Tohkura, 1992 for similar conclusions).

The final noteworthy aspect of Figure 1b is the strong linear correlation between F2 and F3 for the flap distribution. In fact, the correlation coefficient is 0.78 compared to a correlation of 0.17 for English productions (collapsed across phonetic category). This lack of independence in the acoustics may result in a lack of perceptual independence of these dimensions for native Japanese speakers. It is possible that this F2 x F3 dependence may further exacerbate the problem of developing an L2 weighting strategy that requires a re-weighting of F2 and F3.

Japanese Productions of /l-/r/

Figure 1c displays the F2 and F3 onset values for the L2 productions of the Japanese talkers. Figure 1d repeats this display along with the distribution for the L1 flap. We can see here indications of all three potential problems with L2 categorization mentioned in the previous section. First, the decrease of distinctiveness near the English /l-/r/ boundary because of the overlapping flap distribution is evident in the poorly segregated L2 distributions. Whereas the two categories can be perfectly factored by a linear boundary in 1a, there is no such boundary for 1c. Second, there is evidence of perseveration of the L1 weighting strategy. The regression model fit to the L2 productions included beta weights of 0.344 and 0.543 for F2 and F3, respectively (compared to 0.938 and 0.277 for English). That is, there is a greater amount of category label variance accounted for by F2 for Japanese liquid productions than for English. In addition, the relationship of F2 to the two categories is changed in the Japanese productions. For Japanese, a lower F2 is associated with /r/ exemplars. In some respects, the L2 distributions resemble the L1 distributions for the flap and /w/. This is not surprising as Japanese speakers are sometimes taught that /r/ is like /w/ and /l/ is like the flap. These results are also consistent with category or phoneme assimilation accounts such as offered by Flege (1995) and Best (1994). However, the L2 distributions do not line up exactly with the L1 distributions (which would presumably occur with complete assimilation). Instead, The resulting distributions seem to be a compromise between the L1 weighting strategy (high F2 weight) and the optimal L2 weighting strategy (high F3 weight). The fact that the relationship between F2 and category label is reversed for Japanese L2 productions may be a result of the third problem facing the L2 learner mentioned above. That is, the strong correlation of F2 and F3 in L1 distributions may make it difficult to manipulate these two dimensions independently in L2. If this is the case, then the low F3 of /r/ may necessitate a low F2 for Japanese speakers because of interference from experience with L1 phonetic structure. Consistent with this hypothesis, the correlation between F2 and F3 remains high for Japanese productions of the L2 categories at 0.55 (compared to a non-significant 0.17 for English productions).

CONCLUSIONS

The results of this study point to three difficulties for Japanese speakers acquiring the English liquid contrast: 1) overlap of an L1 distribution with the boundary between the L2 distributions; 2) an L1 weighting strategy that is inappropriate for L2; and 3) the lack of independence of two dimensions in L1 that must be varied in L2. We believe that these problems may underlie other L2 acquisition problems and that the solution to these difficulties may also be found in a training approach that is based on a general categorization framework.

ACKNOWLEDGMENTS

The research and preparation of this report were supported by NIH grants 5 R01 DC004674 (A.J.L.) and 5 R01 DC00427 (R.L.D.). Experimental protocols were reviewed and approved by the IRB of the University of Texas.

REFERENCES

- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. V. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167-224). Cambridge, MA: MIT.
- Dalston, R. M. (1975). Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*, *57*, 462-469.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-273). Baltimore: York.
- Goto, H. (1971). Auditory perception by normal Japanese of the sounds L and R. *Neuropsychologia*, *9*, 317-323.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. i., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.
- Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, *39*, 770-784.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman and Company.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331-340.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.
- O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial /w,,j,r,l/ in English. *Word*, *13*, 25-43.
- Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech* (pp. 1-38). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view*. Unpublished Dissertation, U. C.-Irvine.
- Yamada, R. A., & Tohkura, Y. (1992). Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. *Proceedings of ICSLP*, 757-760.