

Acquiring New Speech Sounds by Clustering

Yasuaki Sakamoto (yasu@psy.utexas.edu)

Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA

Abstract

Acquiring a second language phonetic contrast involves learning new sound categories from examples. Consequently, research in category learning should have important implications for research examining how a foreign sound contrast is acquired. The current project is an initial attempt to apply work in category learning to work in second language sound acquisition. Here, I advance that a clustering account developed in category learning research can account for not only various empirical phenomena in the category learning literature but also some critical aspects of acquisition of English /l-/r/ distinction by adults who are native speakers of Japanese surprisingly well. The clustering account suggests new training methods that should aid adults' acquisition of new speech sounds.

In learning to perceive and produce a second language phonetic contrast, learners must acquire new sound categories from examples. Thus, acquiring a foreign sound contrast is an act of category learning. Then, one would expect that advances in our understanding of category learning would play a critical role in understanding how humans acquire new speech categories.

Unfortunately, there has not been a strong link between research in category learning and research in foreign speech sound acquisition. The current project is an initial attempt to make an explicit connection between these two fields by applying what is known in one field to the other. Here, I apply a clustering account developed in category learning research to various aspects of acquisition of English /l-/r/ distinction by adults who are native speakers of Japanese.

Category Learning by Clustering

Research in category learning seeks to better understand how humans encode, organize, and use knowledge. The clustering account of category learning (e.g., Love, Medin, & Gureckis, 2004) posits that humans represent categories by one or more clusters. The clustering model starts with one cluster centered upon the first item encountered, and an item is assigned to the cluster it is most similar to (e.g., closest in multi-dimensional space).

A critical mechanism of the clustering model is its formation of new clusters in response to surprising events or expectation violation (cf. Rescorla & Wagner, 1972). Surprising events can take two forms (e.g., Gureckis & Love, 2003): (a) when people predict an incorrect category (e.g., a child predicts that a bat belongs to the bird cluster, makes a mistake, and develops a new cluster to encode the bat) and (b) when people encounter an item that is not similar to any existing cluster (e.g., one might create a new cluster for pen-

guins, which look quite different from typical birds). Both of these two forms of surprising events may operate in supervised learning, in which corrective feedback is provided (e.g., you are told bats are mammals and not birds). In contrast, in unsupervised learning, in which there is no corrective feedback, only encountering an item that is highly dissimilar to any existing cluster results in a surprising event and the development of a new cluster (Sakamoto & Love, 2004). Using this mechanism, the clustering model can account for various empirical phenomena (Love et al., 2004).

A strong support for the cluster recruitment mechanism is that people better remember items that deviate their expectation (e.g., bats) than items that follow their expectation (von Restorff, 1933). This robust memory phenomenon has been established in various forms under both supervised and unsupervised learning. For example, given a list of items to memorize, people show a memory advantage for an item that differs from others in some way, such as an American city (Austin) in a list of Canadian cities (Vancouver, Toronto, Montréal). Likewise, deviant faces, behaviors, and category members result in enhanced memory (Sakamoto & Love, in press).

In fact, the memory advantage of deviant items is problematic for category learning models that do not accord special status to surprising events (e.g., Sakamoto, Matsuka, & Love, 2004). For example, exemplar models (e.g., Medin & Schaffer, 1978) store individual items without considering what other items are already stored in memory. Prototype models (e.g., Posner & Keele, 1968) average all members of the same category, including items that deviate one's expectation. These models are insensitive to expectation violation and cannot predict enhanced memory for deviant items. In contrast, the clustering model can predict enhanced memory for deviant items by averaging expectation-following items into the same clusters but storing deviant items in their own clusters (Sakamoto & Love, 2004).

Acquiring New Speech Categories

According to the clustering account, learners' ability to acquire a phonetic contrast of a second language is determined by the existing clusters encoding the sounds of their native language. For example, it will be difficult to form new sound clusters if a cluster already exists that encodes a native sound that is highly similar to the newly experienced sounds. Consistent with this idea, a dominant explanation for the difficulty distinguishing between the English /l/ and /r/ by adults who are native speakers of Japanese (e.g., Logan, Lively, & Pisoni, 1991) is that Japanese speakers map both English /l/ and /r/ to the Japanese flap (e.g., Guion, Flege, Akahane-Yamada,

& Pruitt, 2000; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Vallabha & McClelland, submitted). In terms of clustering, instead of distinct English /l/ and /r/ clusters, Japanese has a flap cluster that overlaps with English /l/ and /r/ (Lotto, Sato, & Diehl, 2004), which interferes with Japanese speakers' acquisition of these English sounds.

To distinguish English /l/ and /r/, Japanese speakers need new clusters to encode these sounds. According to the clustering account, encountering highly dissimilar items results in recruitment of new clusters in unsupervised learning. Indeed, Japanese speakers can learn the English /l-/r/ contrast in unsupervised training when they are initially trained on exaggerated examples of English /l/ and /r/ that are highly dissimilar to Japanese flap (McCandliss et al., 2002). In supervised training, Japanese speakers can learn the English /l-/r/ contrast relatively quickly using non-exaggerated sounds (McCandliss et al., 2002) because attempts to cluster these sounds into the Japanese flap cluster lead to prediction errors. Modeling work along this line is currently being carried out (cf. Vallabha & McClelland, submitted). Importantly, this modeling endeavor seeks to unite previous work in category learning and memory with work in speech acquisition.

The clustering account suggests that items that are contrasted with and differentiated from highly similar items can result in more accurate memory (Sakamoto & Love, in press). One way to improve Japanese speakers' acquisition of the English /l-/r/ distinction may be to develop a procedure that focuses people on differentiating the English sounds from the Japanese flap. Such a technique may also assist learners in discovering the dimension that is useful in discriminating the sounds when appropriate contrast sets and training modes are used. Future work along this line is being prepared.

Finally, acquiring categories involves learning which sources of variability are meaningful and which should be ignored or generalized. In this light, theories of category learning are accounts of how humans capture meaningful variation. Training people on a variety of category members is critical for successful category learning (cf. Lotto et al., 2004) and should lead to better understanding of how people acquire new categories.

Acknowledgments

This work was supported by AFOSR FA9550-04-1-0226 and NSF CAREER 0349101. I would like to thank Todd Gureckis, Matt Jones, Brad Love, Momoko Sato, and Marc Tomlinson for helpful discussions with this line of work.

References

Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2724.

Gureckis, T. M., & Love, B. C. (2003). Towards a unified account of supervised and unsupervised learning. *Journal of Experimental and Theoretical Artificial Intelligence*, *15*, 1–14.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first

report. *Journal of the Acoustical Society of America*, *89*, 874–886.

Lotto, A., Sato, M., & Diehl, R. (2004). Mapping the task for the second language learner: the case for Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. C381–C386).

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*, *111*, 309–332.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 89–108.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*, 534–553.

Sakamoto, Y., & Love, B. C. (in press). Vancouver, Toronto, Montréal, Austin: Enhanced oddball memory through differentiation, not isolation. *Psychonomic Bulletin & Review*.

Sakamoto, Y., Matsuka, T., & Love, B. C. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the 6th International Conference of Cognitive Modeling* (pp. 261–266). Mahwah, NJ: Lawrence Erlbaum Associates.

Vallabha, G. K., & McClelland, J. L. (submitted). Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps.

von Restorff, H. (1933). Analyse von vorgängen in spurenfeld. I. Über die wirkung von bereichsbildungen im spurenfeld [Analysis of processes in the memory trace. I. On the effect of group formations on the memory trace]. *Psychologische Forschung*, *18*, 299–342.