Categorization of spectrally complex non-invariant auditory stimuli in a computer game task

Travis Wade & Lori L. Holt
Department of Psychology & Center for the Neural Basis of Cognition
Carnegie Mellon University

ABSTRACT

This study examined perceptual learning of spectrally complex nonspeech auditory categories in an interactive multi-modal training paradigm. Participants played a computer game in which they navigated through a three-dimensional space while responding to animated characters encountered along the way. Characters' appearances in the game correlated with distinctive sound category distributions, exemplars of which repeated each time the characters were encountered. As the game progressed, the speed and difficulty of required tasks increased and characters became harder to identify visually, so quick identification of approaching characters by sound patterns was, although never required or encouraged, of gradually increasing benefit. After thirty minutes of play, participants performed a categorization task, matching sounds to characters. Despite not being informed of audio-visual correlations, participants exhibited reliable learning of these patterns at post-test. Categorization accuracy was related to several measures of game performance and category learning was sensitive to category distribution differences modeling acoustic structures of speech categories. Category knowledge resulting from the game was qualitatively different from that gained from an explicit unsupervised categorization task involving the same stimuli. Results are discussed with respect to information sources and mechanisms involved in acquiring complex, context-dependent auditory categories, including phonetic categories, and to multi-modal statistical learning.

1. Introduction

Underlying objective: understanding how complex auditory categories (including speech sounds) are grouped into equivalence classes based on experience with acoustic structure

Problem: difficulty characterizing previous experience with naturally occurring sounds, particularly speech sounds, hinders understanding of processes contributing to patterns observed in adult (or even infant) perception

~Experience can be controlled for some stimuli in auditory categorization designs

Most current studies: explicit training

Category labels
Discrete response trials
Explicit feedback

Not very much like (for instance) category exposure in language acquisition

Complex correlations between acoustic events sequences and various visual, auditory, olfactory, tactile, and other events in environment.
Learning probably involves implicit sensitivity to these correlations
e.g. Jusczyk, 1997; Kuhl *et al.*, 1992; Maye *et al.*, 2002; Saffran *et al.*, 1996, 1997, 1999; Lacerda and Sundberg, 2004

As a result, explicit training may not involve naturalistic learning mechanisms

2. Method

goals:
~ incorporate both explicit/implicit elements (gradually move from one to the other)
~ involve co-occurrences, only functional labeling
~ stimuli: larger inventory, more difficult (speechlike) categories

Computer game: IRFBATs
(*Incidental robot figure – based auditory training*)
~Simple fps
~Player moves constantly forward
~Periodically approached by animated characters

~4 types of *irfbats*, distinguished by shape, motion and color

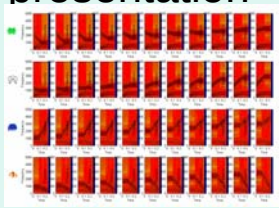
~task: shoot/capture characters appropriately
~accomplished by navigating/aiming, targeting, shooting

~basic idea: high score – by shooting fast
~life decreases when characters are missed; increases when characters caught

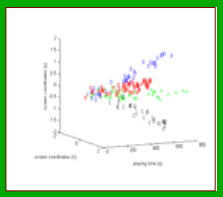
Baptista, L. F. and Petrinovich, L. (**1986**). "Song development in the white-crowned sparrow: social factors and sex differences," Anim. Behav. **34**, 1359-1371.
Bruner, J. (**1983**). *Child's talk: Learning to use language*. New York: Norton.
Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1955**). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. **27**, 769-773.
Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1964**). "Formant transitions and loci as acoustic correlates of place of articulation in American fricatives," Studia Linguistica **18**, 104-121.
Eales, L. A. (**1989**). "The influences of visual and vocal interaction on song learning in zebra finches," Anim. Behav. **37**, 507-508.
Jusczyk, P. W. (**1997**). *The discovery of spoken language*. Cambridge, MA, US: The MIT Press.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by six- months of age," Science **255**, 606-608.
Kuhl, P. K., Tsao, F.-M., and Liu, H.- M. (**2003**). "Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning," Proc. Natl. Acad. Sci. **100**, 9096-9101.
Lacerda, F. (**2003**). "Phonology: an emergent consequence of memory constraints and sensory input," Read. and Writ. An Interdiscip. J. **16**, 41-59.
Lacerda, F., and Sundberg, U. (**2004**). "An ecological theory of language learning," J Acoust Soc Am. **116**, 2523.
Maye, J., Werker, J. F., and Gerken, L. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**, B101-B111.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month-old infants," Science **274**, 1926-1928.
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (**1997**). "Incidental Language Learning: Listening (and Learning) out of the Corner of Your Ear," Psychol. Sci. **8**, 101-105.
Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). "Statistical learning of tone sequences by human infants and adults," Cognition **70**, 27-52.
Sussman, H. M., Hoemeke, K. A., and Ahmed, F. S. (**1993**). "A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation," J. Acoust. Soc. Am. **94**, 1256 - 1268.
Sussman, H. M., Fruchter, D., Hilbert, J., and Sirosh, J. (**1998**). "Linear correlates in the speech signal: The orderly output constraint," Behav. and Brain Sci. **21**, 241-299.

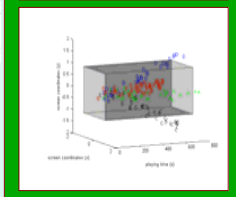
3. Auditory presentation

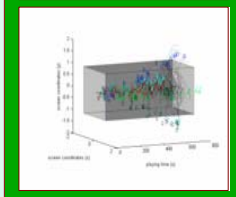
~each character associated with one sound category
~ when character encountered, category exemplar presented
~ sound repeated as long as character is active



~ category (defined by multiple exemplars) tends to co-occur over the game w/ irfbat







~ beginning of game, exposure to patterns more or less implicit:
~ knowledge /attention to sounds is of no apparent consequence in game
~ no mention of importance beforehand
~ additional background score, other sounds

~ however, game becomes progressively difficult:
1) faster
2) characters begin from locations gradually further from center of screen:
~each *irfbat* has characteristic direction
~gradually harder based on sight alone
~eventually, learning sound categories becomes beneficial to game performance

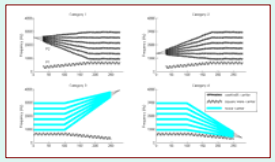
Thus, learning could involve any combination of:

Unsupervised learning from exposure to high-order distributional regularities

"Feedback" involving cross-modal regularities

Functional, task-relevant feedback

4. Stimuli



~ 4 nonspeech categories (6 exemplars each)
~ 250 ms length
~ Created from 2 independent carrier signals, band passed filtered for two spectral peaks (P1, P2)
~ Categories defined by spectrotemporal patterns of P2 (onset or offset)
~ Locus/steady state equation (related to speech production/perception – e.g. Delattre et al., 1955, 1964; Sussman et al., 1993, 1998)
~ Categories 1 and 2 non-invariant in a single dimension, but are when including multiple dimensions

Experiment 1

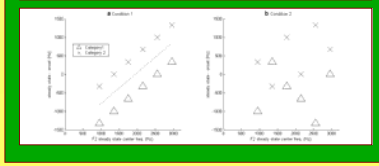
Can people learn categories from game?

Also, is any learning sensitive to differences in category structure?

Procedure: participants played the game for 30 min, took test – matched characters to sounds

Cond. 1: categories possessed higher-order structure

Cond. 2: identical variability in transition/ steady-state space, but no higher-order structure



Experiment 2

How does category awareness resulting from the game compare with explicit training?

Could similar learning result in the absence of the game task?

Procedure: Explicit unsupervised training – participants listened to repeated sound exemplars and grouped them into categories based on salient acoustic characteristics

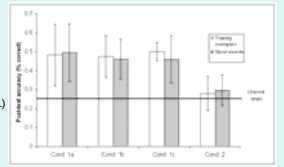
~no feedback

5. Results

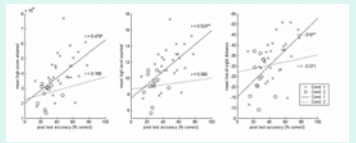
Experiment 1

~ Participants reliably learned sound categories
Condition 1 participants identified sounds above chance ($p<0.001$)

~ But only when they were structured
Condition 2 participants performed at chance, significantly worse than cond. 1 ($p=0.003$)



Category learning:
~occurred
~sensitive to differences in category structure
~related to game performance



Post-test accuracy correlated with players' ability to aim effectively, reach high scores/levels: learning also helped game performance

Seems to indicate interactive process: consistent with processes thought to be important in various types of learning (Baptista and Petrinovich, 1986; Eales, 1989; Bruner, 1983; Kuhl *et al.*, 2003; Lacerda, 2003)

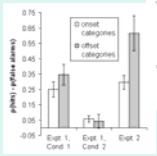
Experiment 2

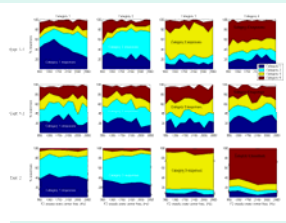
Comparison with explicit learning:

~Explicit task participants also learned, overall slightly more than game participants

~But differences depended on category structure

Category x Training interaction ($p=0.003$) indicated that higher-order structures were learned better as a result of game play





Conclusions

~ categories are learned from game exposure

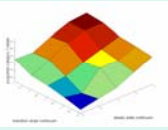
~ learning related to game performance

~ learning sensitive to category structure

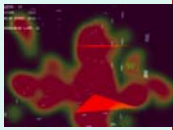
~ learning different from unsupervised task: advantage for higher-dimensional contrasts

Immediate further questions

What is learned about category structure? Will longer-term exposure result in more effective learning, more detailed category knowledge?



What can more sensitive, less explicit measures reveal about category knowledge?



Would similar knowledge result from more extensive explicit training?

How do learned categories relate to naturally acquired categories such as speech sounds?