

Language Acquisition as Complex Category Formation

Andrew J. Lotto

Loyola University Chicago, Ill., USA

Abstract

Purported units of speech, e.g. phonemes or features, are essentially categories. The assignment of phonemic (or phonetic) identity is a process of categorization: potentially discriminable speech sounds are treated in an equivalent manner. Unfortunately the extensive literature on human categorization has typically focused on simple visual categories that are defined by the presence or absence of discrete features. Speech categories are much more complex. They are often defined by continuous values across a variety of imperfectly valid features. In this paper, several kinds of categories are distinguished and studies using human subjects, animal subjects and computational models are presented that endeavor to describe the structure and development of the sort of complex categories underlying speech perception.

Copyright © 2000 S. Karger AG, Basel

There appears to be a tendency in work on speech communication to presume that the form of speech (and language) is a given and that one must hypothesize elaborate mental processes that can accommodate the complexities of this divinely ordained communication system. For example, Chomsky [1957, 1965] endowed children with a specialized Language Acquisition Device to allow them to become competent in the complex recursive rules of language and to discover the underlying structure in the hopelessly impoverished speech of their parents. Liberman and Mattingly [1985] proposed a specialized speech-perception module to help the poor listener deal with the 'lack of invariance' problem that they are unfortunately saddled with because of the variability inherent in speech.

Alternative to this viewpoint, one can presume that the general perceptual and cognitive processes of humans are the givens and that the specific form of our communication system evolved to take advantage of the specific operating characteristics of our cognitive system. I refer to this view as the General Auditory and Learning Approach (GALA) [Lotto, 1996]. It is probably best exemplified in the work of Lindblom [e.g. Liljencrants and Lindblom, 1972; Lindblom et al., 1983; Lindblom, 1986] and Ohala [e.g. 1974, 1999]. GALA is founded on the notion that the development of particular linguistic systems and speech as a communication system is constrained by our general inherited cognitive systems and properties of speech production. In this

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2000 S. Karger AG, Basel
0031-8388/00/0574-0189
\$17.50/0
Accessible online at:
www.karger.com/journals/pho

Andrew J. Lotto
Department of Psychology, Loyola University Chicago
6525 North Sheridan Road, Chicago, IL 60626 (USA)
Tel. +1 773-508-8227, Fax +1 773-508-8713
E-Mail alotto@luc.edu

approach, the characteristics of speech and particular linguistic systems offer opportunities to study mechanisms of perception and cognition.

A good example of GALA is Björn Lindblom's [1986] attempt to predict typical vowel inventories by computing auditory distinctiveness. The accuracy of these predictions is enhanced by more detailed information about the operating characteristics of the peripheral auditory systems. Thus, we see the fingerprints of the auditory system on the content of linguistic sound systems. From these results, it is certainly reasonable to suggest that the contents of vowel systems are constrained in part by a goal of sufficient auditory distinctiveness, where this metric is a function of the capabilities of the auditory system.

This is one example of the many successes of explaining structure and function in speech communication by appealing to constraints of the auditory and articulatory systems. Because of this area of productivity, researchers who eschew the notion of specialized speech mechanisms have sometimes been called 'auditorists' [e.g. Nearey, 1997]. However, this term refers to only one half of GALA. There has always been an assumption in the work of the 'auditorist' that learning plays an important role in accounting for speech behavior. The auditory system provides a representation of the acoustics of a sound, but it is through general learning processes that a listener develops a representation that is useful for communication.

Unfortunately, empirical work on this learning component has been lacking relative to the auditory component. In particular, it is not clear how the characteristics of our general learning processes constrain the form of speech [though, see Lindblom et al., 1983; Nearey, 1997]. What would the fingerprints of the learning system look like? To begin to formulate an answer, let's look at the task for the language learner.

Language Acquisition

The task for infants learning their first language or adults learning a second language appears daunting. Some of the variance in the acoustic input that they receive is directly relevant to the intended message of the speaker. Other variance in the input, however, is the result of extra linguistic influences such as the particular structure of the speaker's vocal tract. The language learner must parse the input variance to discriminate those contrasts that carry information and ignore variation within a contrast that is due to speaker characteristics, coarticulation, articulatory undershoot, etc. Complicating this task is the fact that the language learner must do this in a language-appropriate manner. Languages utilize some subset of over 800 sounds as phonemes and this subset can range from 11 to 141 phonemes [Maddieson, 1984]. As a result of this diversity, variance that is extralinguistic in one language community may be pivotal for discovering the intended message of a speaker in another language environment.

Thus, the task for the language learner is to discriminate some of the acoustic variance and to treat the remaining, potentially discriminable, variance as functionally equivalent. In other words, the language learner must create auditory *categories* that map the linguistically relevant distinctions for the particular language he is attempting to learn. Categorization is a general perceptual process. Much of our perceptual behavior entails treating potentially discriminable stimuli as equivalent. Thus, we may be able to predict the constraints of learning on speech by understanding the operating characteristics of our general categorization processes.

Unfortunately, the empirical work on perceptual categorization has focused primarily on visual categories that are defined by the presence or absence of discrete cues. In contrast, speech sound categories (phonemes) are auditory and are defined by values across a number of imperfectly valid continuous cues. In order to describe language acquisition as categorization and to show the effect of categorization on the structure of speech, we need to understand the processes of complex auditory category formation. My colleagues and I have begun attempts to empirically define the important constructs of learning and categorization to try to supplement our understanding of the role of audition *and* learning in GALA.

Functionally Defined Categories

In order to study language acquisition in speech, we need to develop stimulus sets that contain the same degree of complexity as speech sound categories. Simple tones or clicks will not do. My students and I have tried to create a complex auditory stimulus set that is easily manipulated and that will not be identifiable as speech or any other environmental sound. We believe that we have a set of stimuli that fulfill these desired characteristics. The stimuli are sculpted from 300-ms bursts of white (Gaussian) noise. White noise, as opposed to single tones or collections of several tones, has energy across the range of frequencies; in this way, it is like speech. Using digital signal processing, three attributes are added to the white noise bursts. First, a linear ramp in amplitude defines the onset of the stimulus. This ramp varies in duration (from 10 to 100 ms), which gives the stimuli different degrees of 'attack'. The other two attributes are frequency 'notches' in the noise created by band-stop filters. One 300-Hz-wide notch (where energy is greatly attenuated) varies in low-frequency cutoff from 400 to 850 Hz. The second 300-Hz-wide notch varies from 2,200 to 3,100 Hz. These notches, or gaps, are referred to as NF_1 and NF_2 , respectively (for negative first formant and negative second formant, to make obvious the analogy to vowel stimuli, thus serving a mnemonic purpose despite being an abuse of the term 'formant').

After adding these attributes, the resulting noises sound nothing like speech, but have a desirable amount of complexity. The three attributes can vary independently over distinct continuous ranges. Arbitrary categories can be created across any combination of the three attributes. We had 5 adults learn two categories constructed from this stimulus set across 10 one-hour sessions. Category A contained stimuli that had onsets less than 50 ms, NF_1 s lower than 600 Hz, and NF_2 s greater than 2,800 Hz. Category B was the complement of this set. A stimulus belonged to A if two of the three attributes fell within the range of the description of category A (e.g. onset = 20 ms, NF_1 = 450 Hz, and NF_2 = 2,500 Hz). As compared to typical procedures in categorization, this is an extremely complicated categorization task. The categories are defined on attributes that vary continuously and none of the attributes are necessary nor sufficient to define the category. Listeners must integrate across all attributes and they must do it quickly because they hear the stimulus presented on each trial only once. These characteristics of the task make it very much like the task of learning a speech sound category.

The subjects were presented each sound and asked to press a button labeled A or B. After responding, they received feedback in the form of a light appearing above the correct response button. The question was whether humans could learn to perform this

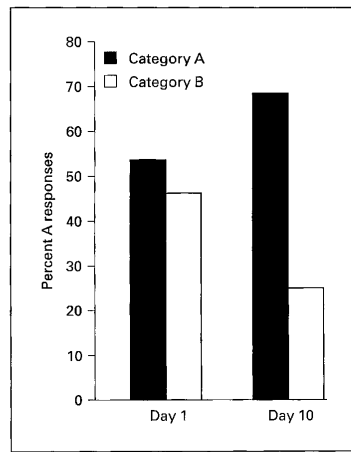


Fig. 1. Percent of category A responses on 1st day and 10th day of training. Filled bars are correct responses (i.e. the sound did come from category A). Unfilled bars are incorrect responses (i.e. sound actually was a member of category B).

complicated task. Despite subjects' concerns about the difficulty of the task, they were able to learn the categories to some extent after less than 10 h. Figure 1 is a graph showing percent A responses after the 1st day of training and after the 10th day of training.

These results demonstrate that auditory categories with the complexity of speech *can* be learned by humans. These novel categories have many similarities to speech sound categories. In particular, these nonspeech categories 'suffer' from the problem of *lack of invariance*. None of the three attributes were necessary nor sufficient to define the category. Yet, subjects were able to learn the categories by simply using general categorization processes. This is similar to the demonstration that birds can learn to correctly identify (categorize) syllables starting with /d/ despite the lack of a single defining cue [Kluender et al., 1987].

In this experiment, the categories were defined completely by the response required for the stimuli (as indicated by feedback). That is, stimuli that required an equivalent response were considered members of the same category. I refer to these as functionally defined categories. Speech sound categories are certainly functionally defined to some degree. That is, sounds are grouped into a phonemic category because they are functionally equivalent when it comes to lexical access. This kind of *functional-equivalence* category is similar to traditional definitions of linguistic categories. For example, phonemes are often defined in terms such as: '... a family of uttered sounds... in a particular language which count for practical purposes as if they were one and the same...' [Jones, 1967, p. 258]. However, there is also systematic variance in the input distributions of speech sounds that can serve as information for a language learner about the definition of phonemic categories. For example, the distributions of voice onset time for voiced and voiceless consonants in various languages are Gaussian-like in shape with exemplars in the middle of a category having higher frequencies of occurrence than exemplars near the boundaries between categories [Lisker and

Abramson, 1964]. This statistical information could be used by language learners to parse the space of voice onset time into phonemic categories. In addition, correlations between features can provide information about categories, as correlations tend to be higher within a natural category than between categories.

Statistical information must be playing some role in language acquisition, because infants show evidence of native-language phonemic categorization at 6 months of age before a lexicon is established to define functional equivalence [Kuhl et al., 1992]. The input distributions for the functionally defined nonspeech categories presented to our subjects in the experiment described above were rectangular with no correlational structure. This may be why our subject found these categories rather difficult to learn. What happens if we add statistical information? Does this change the type of learning involved?

Functional-Statistical Category

As mentioned above, speech-sound categories are good examples of categories that are functionally and statistically defined. Of course, it is difficult to study the formation of speech-sound categories in humans because it is difficult to ethically control input or to get a valid measurement of what input a child or second-language learner actually receives. However, one *can* control precisely the input to a nonhuman animal and, thus, study the response structures arising from functional-statistical category formation (with the presumption that animal and human perceptual categorization processes are fundamentally similar). My colleagues at the University of Wisconsin and I have run such a study [Kluender et al., 1998].

Birds (European starlings, *Sturnus vulgaris*) were trained to peck in response to exemplars from one vowel category (e.g. /i/) and to refrain from pecking when presented exemplars from a second category (e.g. /ɪ/). The exemplars were chosen from stylized distributions and varied in first (F_1) and second (F_2) formant frequencies. The birds were reinforced for pecking to exemplars from one of these distributions (the positive response category was randomly assigned to each bird). That is, the categories were defined functionally; all sounds in the /i/ category were to be responded to equivalently (by pecking the key). In addition, there was statistical information about category boundaries in the input distributions during training. The vowel distributions were nonoverlapping. The area around the centroid of each distribution (in $F_1 \times F_2$ space) was more densely sampled (though the true centroid of the distribution was not presented during training) and fewer exemplars were sampled from the boundaries of each distribution. Thus, one could detect the categories by differentiating input density across $F_1 \times F_2$ space. Could this information affect avian categorization of vowel distributions?

After less than 100 h of exposure to this task, the birds were tested on the categories. They demonstrated excellent categorical behavior. Peck rates to positive vowels (e.g. /i/) were substantially higher than pecks to negative vowels (e.g. /ɪ/). The birds easily generalized their responses to novel exemplars from the categories (e.g. pecking to an /i/ exemplar with F_1 and F_2 values that were never presented before). Because of control conditions within the experimental design, we could determine that birds' responses were due solely to their experience with the sounds and reinforcement conditions. Certain pairs of stimuli were part of a single category for some birds, but straddled two categories for other birds. Starlings for whom the pair members fell within

distinct categories pecked differentially to the pair members, indicating that they were able to discriminate the two. Birds who had been trained to equate the exact same stimuli responded to them equivalently.

The birds also showed a gradient in their response structures. They pecked far more vigorously to positive stimuli that were maximally separated in the $F_1 \times F_2$ space from negative stimuli (e.g. stimuli with high F_2 and low F_1 if the positive vowel was /i/) than to stimuli that were near the boundary with the other category. This gradient in response may be a general consequence of learning a functional equivalence class. Classic work in discrimination learning describes results in which responses to a positive stimulus tend to strengthen as one moves away from the negative stimulus [e.g. Spence, 1936, 1937, 1952, 1960; Hansen, 1959; Mackintosh, 1995]. Thus, a response gradient may be a fingerprint of the processes involved in forming functionally defined perceptual categories. Do we see this gradient in human speech categories?

Additional data from this project demonstrates that a similar structure is apparent in responses of humans judging the representativeness or 'goodness' of vowels [Kluender et al., 1998]. We presented human adults with the same vowel distributions that were presented to the birds. Human listeners judged the stimuli in terms of 'goodness' as exemplars of the English vowels /i/ and /ɪ/. The responses of the adults showed a gradient similar to that exhibited by the birds. The best /i/ exemplars were judged to be those furthest from the /ɪ/ distribution (i.e. low F_1 and high F_2). Interestingly, these exemplars of /i/ would be very rare in natural speech because vowels are often reduced (moved away from the extremes of the $F_1 \times F_2$ space) in normal speaking contexts [Lindblom, 1963; Johnson et al., 1993]. Other researchers have found similar gradients in human perceptual responses to vowels [Johnson et al., 1993; Aaltonen et al., 1997; Lively and Pisoni, 1997]. In all cases, listeners appear to prefer vowels that are maximally distinguished from competing vowel categories. Thus, this gradient may be a nice example of the fingerprint of general mechanisms underlying learning of functionally defined categories.

Besides this gradient, there was a second salient feature present in the structure of birds' pecking responses. Birds tended to peck more to the exemplar at the centroid of the positive distribution than to more remote exemplars. (This resulted when averaging across all stimulus tokens of the distribution. On a token-by-token basis, the highest peck rates were not at the centroid, but at the extremes of the space. This latter finding defines the gradient as described above.) This is despite the fact that the bird had never experienced the centroid stimulus during training. This pattern of responses is similar to patterns that have been used as justification for *prototype* models of classification in the general categorization literature [Posner and Keele, 1968; Rosch, 1973, 1988]. This prototype structure is interesting because it demonstrates that the birds learned something about the structure of input distributions. The centroid stimulus came from the most densely sampled region (in $F_1 \times F_2$ space) of the input distribution. Birds' responses reflected this statistical fact of the distributions, although it was not necessary for the birds to learn about the statistical structure of the input to perform the task correctly. The functional equivalence classes that the birds were to learn could be defined by a simple (linear) boundary in stimulus space. Nonetheless, there was evidence that birds' responses were affected by statistics of the input structure; the birds seemed to pick up statistical information about the task incidentally. This is consonant with many recent findings demonstrating an amazing ability for adult and infant humans to learn statistical information (e.g. transitional probabilities) in auditory

streams, even when the information is unrelated to any current task [Saffran et al., 1996, 1997, 1999].

Thus, it may be that the prototype in the response structure is a fingerprint of the general processes underlying the formation of statistically defined categories. Do we see anything like this in human speech sound categories? Yes. Similar prototype structures were also present in the 'goodness' ratings of our human adult subjects for the /t/ vowel distributions used in the avian learning study. In fact, the agreement between the birds' responses and the humans' judgments was quite amazing. The correlation coefficient across vowels was $r = 0.99$ and the average r within any particular vowel distributions was about 0.7. Both humans and birds demonstrated gradients (indicative of functionally defined categories) and prototypes (indicative of statistically defined categories). We were also able to model these response structures fairly successfully with a simple linear associator (conceptualized as a neural network). Together these data suggest that general categorization processes may play a role in the development and maintenance of phonetic categories.

Conclusions

These previous experiments serve as initial attempts to discover the kinds of effects that general learning processes may have on the form of speech categories. They have demonstrated that *prototype* effects and solutions to the *lack of invariance* problem may be delivered by general processes of categorization. Additionally, recent work on computational models of learning suggests that decreased intracategory discrimination or *categorical perception* is an expected result of any categorization process [Dampier and Harnad, in press]. One may note that *lack of invariance*, *prototypes*, and *categorical perception* are all concepts that one time or another have been proposed as evidence for specialized mechanisms for speech perception. Research on general categorization processes offers hope of a parsimonious explanation for all these phenomena. With continuing work on the systems of learning and audition, GALA now holds promise as a coherent and integrated framework for understanding structure and function in speech.

References

- Aaltonen, O.; Eerola, O.; Hellström, Å.; Uusipaikka, E.; Heikki Lang, A.: Perceptual magnet effect in the light of behavioral and psychophysiological data. *J. acoust. Soc. Am.* 101: 1090–1105 (1997).
- Chomsky, N.A.: Syntactic structures (Mouton, The Hague 1957).
- Chomsky, N.A.: Aspects of the theory of syntax (MIT Press, Cambridge 1965).
- Dampier, R.I.; Harnad, S.R.: Neural network models of categorical perception. *Percept. Psychophys.* (in press).
- Hansen, H.M.: Effects of discrimination training on stimulus generalization. *J. exp. Psychol.* 58: 321–372 (1959).
- Johnson, K.; Flemming, E.; Wright, R.: The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69: 505–528 (1993).
- Jones, D.J.: The phoneme (Cambridge University Press, Cambridge 1967).
- Kluender, K.R.; Diehl, R.L.; Killeen, P.R.: Japanese quail can learn phonetic categories. *Science* 237: 1195–1197 (1987).
- Kluender, K.R.; Lotto, A.J.; Holt, L.L.; Bloedel, S.B.: Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models. *J. acoust. Soc. Am.* 104: 3568–3582 (1998).
- Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K.N.; Lindblom, B.: Linguistic experiences alters phonetic perception in infants by 6 months of age. *Science* 255: 606–608 (1992).
- Lieberman, A.M.; Mattingly, I.G.: The motor theory of speech perception revised. *Cognition* 21: 1–36 (1985).

- Liljencrants, J.; Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: 839–862 (1972).
- Lindblom, B.: Spectrographic study of vowel reduction. *J. acoust. Soc. Am.* 35: 1773–1781 (1963).
- Lindblom, B.: Phonetic universals in vowel systems; in Ohala, Jaeger, *Experimental phonology* (Academic Press, Orlando 1986).
- Lindblom, B.; MacNeilage, P.; Studdert-Kennedy, M.: Self-organizing processing and the explanation of phonological universals; in Butterworth, Comrie, Dahl, *Explanations of the phonetic universals* (Mouton, The Hague 1983).
- Lisker, L.; Abramson, A.S.: A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20: 384–422 (1964).
- Lively, S.E.; Pisoni, D.B.: On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *J. exp. Psychol. hum. Perception Performance* 23: 1665–1679 (1997).
- Lotto, A.J.: General auditory constraints in speech perception: the case of perceptual contrast; PhD diss. University of Wisconsin-Madison (1996).
- Mackintosh, N.J.: Categorization by people and pigeons: the twenty-second Bartlett Memorial Lecture. *Q. Jl. exp. Psychol.* 48: 193–214 (1995).
- Maddieson, I.: *Patterns of sound* (Cambridge University Press, Cambridge 1984).
- Nearey, T.M.: Speech perception as pattern recognition. *J. acoust. Soc. Am.* 101: 3241–3254 (1997).
- Ohala, J.J.: Experimental historical phonology; in Anderson, Jones, *Historical linguistics II: Theory and description in phonology* (North Holland, Amsterdam 1974).
- Ohala, J.J.: Acoustic-auditory aspects of speech. 35th Regional Meet. Chicago Ling. Soc., Univ. Chicago, April 1999.
- Posner, M.I.; Keele, S.W.: On the genesis of abstract ideas. *J. exp. Psychol.* 77: 353–363 (1968).
- Rosch, E.: Principles of categorization; in Collins, Smith, *Readings in cognitive science: a perspective from psychology and artificial intelligence* (Morgan Kaufmann, San Mateo 1998).
- Rosch, E.H.: Natural categories. *Cognitive Psychol.* 4: 328–350 (1973).
- Saffran, J.R.; Aslin, R.N.; Newport, E.L.: Statistical learning by 8-month-olds. *Science* 274: 1926–1928 (1996).
- Saffran, J.R.; Johnson, E.K.; Aslin, R.N.; Newport, E.L.: Statistical learning of tone sequences by human infants and adults. *Cognition* 70: 27–52 (1999).
- Saffran, J.R.; Newport, E.L.; Aslin, R.N.; Tunick, R.A.; Barrueco, S.: Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* 8: 101–105 (1997).
- Spence, K.W.: The nature of discrimination learning in animals. *Psychol. Rev.* 43: 427–449 (1936).
- Spence, K.W.: The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.* 44: 430–444 (1937).
- Spence, K.W.: The nature of the response in discrimination learning. *Psychol. Rev.* 59: 89–93 (1952).
- Spence, K.W.: *Behavior theory and learning* (Prentice-Hall, Englewood Cliffs 1960).