ABSTRACT

Categorical Perception as an Emergent Feature of General Perception

Dale Price

Director: Charles Weaver, Ph.D.

Humans perceive linguistic phonemes categorically, that is, in distinct clusters rather than as continua. The same pattern of categorical perception (CP) has also been documented with a wide variety of other stimuli, including nonspeech sounds and faces. Although CP has often been assumed to be a distinct mode of perception, some studies have suggested that it is an emergent phenomenon that can occur with any general perceptual system. In this thesis I provide an overview of the research on CP, then examine findings on the factors which cause CP in humans and synthetic CP in simulated neural networks in order to address the hypothesis that categorical perception is merely a feature of general perception and suggest possibilities for future investigation.

APPROVED BY DIRECTOR OF HONORS THESIS:

Dr. Charles Weaver, Department of Psychology and Neuroscience

APPROVED BY THE HONORS PROGRAM:

Dr. Andrew Wisely, Director

DATE:

CATEGORICAL PERCEPTION AS AN EMERGENT FEATURE

OF GENERAL PERCEPTION

A Thesis Submitted to the Faculty of

Baylor University

In Partial Fulfillment of the Requirements for the

Honors Program

By

Dale Price

Waco, Texas

December 2013

TABLE OF CONTENTS

Acknowledgments iii

Chapter One: Introduction 1

Chapter Two: Defining Categorical Perception 4

Chapter Three: Experimentation Paradigms 7

Chapter Four: Evidence for CP as an Emergent Phenomenon....................................12

References 19

ACKNOWLEDGMENTS

I would like to thank Dr. Weaver for his time spent as my mentor for this project. I would also like to thank Dr. Achor, Dr. Butler, and Dr. Riley for serving on my thesis committee.

CHAPTER ONE

Introduction

The human mind must interpret varied stimuli from many different sources. Often, this interpretation involves associating an incoming stimulus with similar, already learned stimuli. For example, a person listening to another person speaking associates the sounds of speech with the words he/she already knows and is thus able to discern the speaker's meaning from those words. At a phonemic level, the listener must interpret the raw sounds in terms of the phonemes—basic sound building blocks—of his/her language. Of course, a single phoneme will be acoustically different when pronounced by different people—for example, the different sounds of a male and a female voice, or regional variation of accents. Even the same phoneme pronounced by the same person will be acoustically quite different: the "t" in "today" is not exactly the same as the "t" in "not," and any number of factors might muffle or alter the sound. Nevertheless, anybody fluent in the language can usually discern which speech sounds they hear with little difficulty.

Listeners are able to understand a vast number of different sounds because they perceive them according to the limited set of phoneme categories determined by their language. Infants learning their first language acquire these categories—and the enhanced ability to distinguish them—at approximately one year of age (Werker, Gilbert, Humphrey, & Tees, 1981). However, they also lose the ability to perceive sound distinctions that are not used as phonemes in their language.

Liberman, Harris, Hoffman, and Griffith (1957) described this effect as an acquired similarity for sounds in the same category, and acquired distinctiveness for sounds in different categories. That is, sounds that are habitually given the same phoneme label appear to be more similar than they actually are; for example, a “b” sound in a male voice and a “b” sound in a female voice seem subjectively more similar than their acoustic waveforms are. Similarly, an English speaker hearing an unaspirated “b” sound and an aspirated “bh” sound, which are completely separate phonemes in Hindustani but not in English, may not be able to discriminate between the two sounds as easily as a native Hindustani speaker. On the other hand, sounds habitually given different labels appear more different than they are in reality. The “b” sound and “p” sound produced by the same voice differ only a small amount acoustically by voice onset time (VOT) but sound very different to someone whose language uses those sounds. Later experiments demonstrated a similar categorical effect for nonspeech stimuli in multiple sensory modalities (Levin & Beale, 2000; Locke & Kellar, 1973; Pastore, Li, & Layer, 1990).

Originally, Liberman et al. (1957) predicted that categorical perception (CP) would entail the equivalence of category labeling and discrimination. CP as demonstrated by experimentation has never met that strict definition, but certain experimental paradigms have consistently produced varying degrees of CP in which discrimination was correlated to, but slightly better than, predictions based on labeling. Discrimination in these experimental tasks may involve a mixture of both categorical and non-categorical inputs. Neural network models of perception using simple learning algorithms have demonstrated human-like CP without the need for any innate categorization mechanisms.

CP in humans may be explained as an emergent phenomenon resulting from the interaction between stimuli and general perceptual learning systems. Only learning in the later stages of perceptual processing seems to be necessary to produce CP. However, lower-level perceptual systems may influence certain aspects of CP by means of innate or learned mechanisms. Future research into CP should investigate the precise role different levels of perceptual processing play and the extent to which their effects are innate or acquired in different sensory modalities. Because emergent CP requires environmental factors, research should also investigate what factors enable CP in everyday speech perception outside of the typical CP tasks.

CHAPTER TWO

Defining Categorical Perception

As Schouten, Gerrits, and Van Hessen (2003) noted, various researchers have defined CP differently. Broadly speaking, CP is defined as occurring when between-category discrimination is greater than within-category discrimination. That is, either differences between stimuli in the same category are dulled, those between stimuli in different categories are enhanced, or both, as Liberman et al. (1957) suggested. Classically, CP has been defined based on four features: a sharp category boundary, a discrimination peak corresponding to the boundary, the discrimination function predictable from identification, and resistance to contextual effects (Treisman, Faulkner, Naish, & Rosner, 1995). In light of the finding that response bias may influence the appearance of CP (Schouten et al., 2003), it is necessary to produce an experimental definition of CP that includes the relevant aspects of the phenomenon but excludes the effects of experimental factors and response bias.

Allowing for Partial CP

Liberman et al. (1957) predicted that participants would only be able to discriminate between stimuli that they placed in different categories. In an ABX procedure with synthesized speech sounds, where participants attempted to determine whether the third sound in a triad was identical to the first or second sound, participants discriminated better between sounds to which they attached different phonemic labels than between those in the same category. The assumption that they would only be able to discriminate between stimuli given different labels predicted the high and low points of their discrimination curves; however, the participants performed above the levels predicted by that assumption in general. Liberman et al. (1957) speculated that the better-than-expected performance may have been due to flaws in the recording and playback techniques used in the experiment. Damper and Harnad (2000) later noted that all empirical tests had resulted in discrimination performance slightly above what was predicted by identification, suggesting that it was not due to those flaws in the original experiment.

Instead of the strict definition of CP by Liberman et al. (1957), more recent studies such as Damper and Harnad (2000), Levin and Beale (2000), and Schouten et al. (2003) acknowledged that CP can occur in varying degrees. Listeners may use a mixture of both categorical and non-categorical information. To quantify the degree of CP in experiments, Van Hessen and Schouten (1999) developed the CP index, which expresses degree of CP from 0 to 100 as a function of the correlation between identification and discrimination over the average distance between the two.

Signal Detection Theory in Modern Approaches to CP

Van Hessen and Schouten's (1999) measure of CP index is compatible with a signal detection theory (SDT)-based approach to CP. Schouten et al. (2003) and Gerrits and Schouten (2004) noted that many testing modalities introduce a strong response bias which is difficult to separate from actual CP. An approach using SDT allows for the separation of sensitivity (d') from response bias (β). Damper and Harnad (2000) noted that this enables researchers to distinguish true CP at the sensory perception level from categorization at the later decision level, which would fall under response bias and is not a notable phenomenon. Because they only considered CP at the sensory perception level, they argued that CP can be defined concisely as occurring when identification d' is equal to discrimination d'. This SDT-based definition is much clearer and more testable than classical feature-based definition. The identification and classification d' measures can be used to calculate CP index (Van Hessen & Schouten, 1999), allowing varying degrees of CP to be quantified.

CHAPTER THREE

Experimentation Paradigms

Overview

Different discrimination tasks used in CP experiments allow participants to use auditory and phonetic processing to varying degrees, affecting the degree of CP. The most common task, ABX, presents listeners with stimulus triads: the first two stimuli are different, followed by a third that is identical to one of the first two. The listener attempts to determine which of the first two stimuli matches the third. As in the original experiment by Liberman et al. (1957), CP occurs when listeners perform better for stimuli to which they assigned different labels than for those given the same label. The ABX task was thought to result in a high degree of CP because the limitations of auditory memory would only allow listeners access to phonetic information during the task (Massaro and Cohen, 1983). However, Gerrits and Schouten (2004) noted that the ABX task caused a strong response bias toward the B response, obscuring the separation between sensitivity and bias.

A variant, AX, could reduce the auditory memory requirement—and thus potentially allow the use of auditory in addition to phonetic information—but would likely induce a strong response bias. Another variant, AXB, yielded inconsistent results. Van Hessen and Schouten (1999) found that subjects treated it like an AX task, while Gerrits (2001) reported differences between AX and AXB results.

In the two-interval two-alternative forced-choice (2I2AFC) task, two different stimuli are presented in AB or BA order and the listener attempts to determine which order has been presented. Gerrits and Schouten (2004) note that although the 2I2AFC task involves a smaller amount of response bias than ABX, it may encourage conscious category labeling because the instructions explicitly mention phoneme categories.

The 4IAX task involves two pairs of stimuli: one in which both stimuli are identical, and one in which they are different (Pisoni, 1975). The listener attempts to choose the pair in which the stimuli are different. This task allows listeners to make greater use of auditory cues rather than subjective categories, reducing the amount of CP. A variant, 4I2AFC, was expected to allow listeners either a 4IAX-like strategy—enabling auditory processing—or a 2I2AFC-like strategy—enabling the use of phonetic labels (Gerrits and Schouten, 2004). However, in Gerrits and Schouten's (2004) experiment, participants used only 4IAX-like auditory processing.

Effects of Testing Methodology on CP

The degree of CP produced in various experiments generally depended on the degree to which the task allowed the use of subjective criteria and category labels. Despite the significant response bias in many of the tasks allowing CP, this may support the idea that CP is an emergent feature that requires interaction between certain environmental characteristics and the general perceptual system.

CP Does Not Require Special Mechanisms

Pastore et al. (1990) suggested that the same general perceptual mechanisms may be responsible for both speech and nonspeech CP given an appropriate task. In a study by Mattingly, Liberman, Syrdal, and Halwes (1971), the same sounds were perceived categorically in a speech context and non-categorically in a nonspeech context. Those results do not necessarily indicate separate mechanisms for speech and nonspeech perception; Pastore et al. (1990) hypothesized that subjects in Mattingly et al. (1971) may have simply failed to focus on the critical components of the continuum or learned to respond based on their perceptions of those components. Pastore et al. (1990) employed an ABX procedure using sinusoid-based chirps and bleats labeled “bleat,” “short bleat,” and “chirp.” Participants underwent several sessions of a labeling procedure before the ABX trials to ensure that they had learned to associate the critical components of the continuum with the response categories. Their perception was not perfectly categorical, but they were able to refer to the exemplar categories as a basis for their responses, contrary to the findings of Mattingly et al. (1971).

These results do not definitively indicate that the same general perceptual mechanisms are responsible for CP in Pastore et al. (1990) and phoneme categorization in everyday speech perception. Schouten et al. (2003) suggested that listeners’ use of internal, subjective criteria in a biasing task like ABX, which is not necessarily similar to everyday speech perception, may have allowed them to perceive stimuli categorically. However, the findings of Pastore et al. (1990) do suggest the possibility that CP in both speech and nonspeech perception may be an emergent property of the general perceptual system rather than a special mode.

The results of Levin and Beale (2000) may also suggest that the ability to use certain subjective criteria in the discrimination task enables some degree of CP, even for unfamiliar stimuli. Levin and Beale (2000) tested for CP on novel continua of artificial faces, using the endpoint faces as exemplars for the categories. To eliminate any effects of configural coding, they used inverted faces for one experiment. Participants displayed CP even before they had gained practice in the discrimination trials. With inverted faces, which disrupted configural coding, CP was present but slightly reduced compared to regular faces, suggesting that configural coding is not strictly necessary for CP to occur, but can increase the degree of CP. These results indicate that there are multiple factors which influence the degree of CP—in these experiments, configural coding and the subjective associations enabled by the task—suggesting that CP is not the result of a single, rigid mechanism but of the interaction between several testing factors and general perceptual mechanisms. It may be useful to further investigate the effects of configural coding on CP in a less biasing task without the explicit use of stimulus category labels as in Levin and Beale (2000).

CP Depends on the Testing Methodology

Schouten et al. (2003) noted that in ABX and 2IFC tasks, participants' use of a category labeling strategy led to response bias for stimuli on either side of the phoneme boundary. In a 2IFC task with flanking stimuli, similar to 4I2AFC, participants used a bias-free 4IAX-like strategy, producing no CP—there was no relationship between classification and discrimination along a vowel continuum. In the strongly biasing task without flanking stimuli, participants showed CP—a clear positive relationship between classification and discrimination. Schouten et al. (2003) argued that these results indicate a subjective categorization mechanism enabled by bias in the task, separate from the mechanisms responsible for everyday speech perception. However, the same tasks had also enabled CP in a simulated neural network model with no innate categorization facilities in a study by Damper and Harnad (2000). The presence of CP may be due to some other aspect of the task, regardless of the listener's use of subjective categorization criteria. Nevertheless, it remains unclear whether everyday speech perception would trigger CP in general perceptual processing in a similar way to these tasks.

In Gerrits and Schouten's (2004) experiments, listeners produced a high degree of CP in a 2I2AFC task, consistent with the expectation that the task would allow only a phonemic strategy. They did not demonstrate CP in a 4I2AFC task despite the expectation that it would allow both phonemic and auditory processing. Increasing the interstimulus interval (ISI) gave listeners more time to categorize stimuli while still allowing for auditory processing, but only caused a minimal increase in the degree of CP. Therefore, the time it takes listeners to categorize stimuli was not the factor which prevented them from using category information in this task. However, it is unclear what factors did cause the listeners to use only auditory information in discriminating speech sounds when both auditory and phonetic information should have been available.

CHAPTER FOUR

Evidence for CP as an Emergent Phenomenon

Neural Network Models of Perception

Neural network models of CP have demonstrated that no innate categorization mechanisms are required for CP to occur. In a CP study by Anderson, Silverstein, Ritz, & Jones (1977), a trainable neural network model produced identification and discrimination curves similar to those of human and animal subjects. As in humans and other animals, the synthetic CP of the neural net did not meet the strictest definition of CP—that identification and discrimination were equal—but the model did show some degree of CP. The synthetic CP met the classical definition identified by Treisman et al. (1995): a sharp category boundary with a discrimination peak at the boundary, some degree of predictability of discrimination from identification, and resistance to contextual effects.

Damper and Harnad (2000) applied the brain-state-in-a-box (BSB) neural net model developed by Anderson et al. (1977) and multilayer perceptrons (MLPs) trained on Rumelhart, Hinton, and Williams' (1986) error back-propagation algorithm to model CP of synthesized stop consonants. They used Pont and Damper's (1991) model of the peripheral auditory system to preprocess the synthesized consonant-vowel syllables.

The peripheral auditory processing model employed by Damper and Harnad (2000) simulated any nonlinearities introduced by human auditory transduction, which were thought to contribute to CP. The neural net models were trained using the output of the auditory model, a neurogram depicting the state of auditory nerve fibers in response to each stimulus.

The BSB model is an associative neural net, i.e. it takes a noisy pattern as input and outputs the corresponding prototypical, noise-free pattern. To accommodate CP when creating the model, Anderson, Silverstein, et al. (1977) added positive feedback from the input neurons, and limited the feedback-induced activation by allowing the simulated neurons to become saturated. Consequently, the model could reach a stable state with regard to the connection between inputs and prototypical outputs. For the simulated speech in the experiment by Damper and Harnad (2000), the outputs corresponded to the presence or absence of distinctive features such as consonant voicing. Given the speech neurograms as input, the model produced classical CP including a steep labeling curve, an ABX discrimination peak at the category boundary, and a shift with place of articulation similar to results from human subjects.

Damper and Harnad (2000) also tested the CP of a back-propagation feedforward neural net. Prior research showed that feedforward auto-associative neural nets can perform a principal component analysis of the input. To determine the presence of a CP effect, the discrimination function produced after categorization training is compared to the precategorization discrimination function produced by auto-association training alone (Harnad, Hanson, & Lubin, 1991). Harnad, et al. (1995) identified three important factors in producing synthetic CP in a feedforward net: maximal interstimulus separation during auto-association learning; stimulus movement to achieve linear separability during categorization learning, promoting within-category compression and between-category separation; and inverse-distance repulsion at the category boundary, causing the hidden-unit (MLP) representation to move away from the category boundary. As with the BSB model, Damper and Harnad (2000) trained the neural net on the endpoints with minimal and maximal VOT, then tested the model's generalization using the range of VOT stimuli. The model produced classical CP including a steep labeling curve and a shift of the curve with varying place of articulation. The category boundary values closely matched those in humans. The back-propagation model produced more convincing results than did the BSB model on a simulated ABX experiment because the back-propagation model's discrimination was higher than that predicted mathematically, as in humans.

Damper and Harnad (2000) made a distinction between CP for novel or artificial stimuli and CP for stop consonants. For novel and artificial stimuli, the model simply placed the category boundary in the middle of the continuum or at the label boundaries provided in training, modeling learned categorization in humans. For stop consonants, the model placed the boundaries in roughly the same places as human listeners. Because the BSB and MLP models exhibited these two different behaviors for artificial and VOT stimuli, Damper and Harnad (2000) concluded that the VOT continuum itself must have some special property that creates the potential for CP in a connectionist model. The auditory preprocessor was responsible in some way for the boundary-movement phenomenon, and may have also altered the input in such a way as to support CP. These results demonstrate that any connectionist model of general learning has the potential to exhibit human-like classical CP given inputs that lend themselves to categorization in some way. The model does not need to have any innate programming or special parameters to facilitate categorization. Instead, CP is an emergent phenomenon originating from the interaction between the learning system, auditory perception, and the stimuli.

Damper and Harnad (2000) acknowledged that their models did not simulate memory effects and suggested that future work include recurrent feedback connections to simulate a memory buffer. Also, they were only able to test their models on one type of stimulus, stop consonants, in an ABX paradigm that may not be representative of natural speech processing according to Schouten et al. (2003).

Damper and Harnad (2000) reported that the peripheral auditory model was responsible for the boundary-movement phenomenon, but it is not entirely clear what role the peripheral auditory system model played in altering or facilitating other aspects of CP. Nevertheless, it is clear that transduction does affect CP during the subsequent perceptual processing. Human perceptual systems involve multiple hierarchical stages of processing; the effects of learning in each stage may be involved in producing CP in subsequent stages.

Hierarchical Processing and CP

In visual processing, the task being performed affects the perceptual analysis in the initial stages of visual processing (Sowden & Schyns, 2006). Casey and Sowden (2012) used a neural net model to investigate how neural processing changes at different perceptual levels to facilitate CP. They used a model involving layers of interconnected neurons trained by means of Hebbian adaptation with lateral inhibition through competition. This type of model enables modularity so that different stages of visual processing can be modeled, and learning takes place through association like other neural net models.

Casey and Sowden (2012) modeled key stages in human visual processing using pairs of contralateral modules: pre-cortical, early visual cortical, and ventral visual processing. The model formed feature detectors using competitive learning based on Rumelhart and Zipser's (1985) algorithm. Task influence for categorization was modeled by biasing learning for a category signal in the ventral visual module which was active during training for items in a particular category. Casey and Sowden (2012) trained the models using inputs representing images of compound Gaussian windowed gratings with varying $3f$ phase.

CP was produced at only the highest levels of processing when the stimuli were all in the same orientation (Casey & Sowden, 2012). However, when stimuli in different orientations were used, CP lacked human-like specificity for stimulus orientation. In humans, however, a larger degree of orientation specificity was found. Thus, synthetic CP could be produced at high levels of perceptual processing, but without changes to processing at lower levels it lacked certain human-like aspects.

Future Directions for Investigation

The lack of human-like specificity for high-level-only CP in Casey and Sowden's (2012) experiment reflects Damper and Harnad's (2000) finding that transformations performed by the low-level auditory preprocessor were responsible for the human-like distribution of category boundaries despite not appearing entirely necessary for CP. However, Damper and Harnad's early auditory preprocessor modeled innate transformations to the stimulus during transduction, while any early changes in a model similar to Casey and Sowden's (2012) would be learned based on inter-module feedback. It may be worth investigating to what extent both learned and/or innate features affect CP in different sensory modalities.

Gerrits and Schouten (2004) were not able to determine exactly what factors in the 4I2AFC experiment caused listeners to use only auditory rather than phonetic information, compared to the 2I2AFC task. A neural net model of auditory perception may enable future researchers to investigate why different categorization tasks allow or preclude CP because such a model would allow a more detailed examination of the task's effects on processing. As Damper and Harnad (2000) noted, their model of auditory perception did not simulate time and memory. Massaro and Cohen (1983) speculated that memory effects could play some role in determining what types of information were available to the listener in different tasks, so a neural net model incorporating memory effects could test this possibility.

Schouten et al. (2003) argued that the presence of response bias in traditional discrimination tasks only enabled CP because the tasks caused participants to use internal, subjective criteria. Therefore, they argued, such tasks did not represent natural speech perception. The findings of neural net studies indicate that the CP produced by those tasks is an emergent property of any perceptual learning system, i.e. it does not require any special mechanism but is created by the interaction of the task and the general perceptual system. It is therefore plausible that natural speech perception also engages emergent CP in a similar way to the traditional discrimination tasks, even though it may not be similar to those tasks. Because neural net modeling enables closer examination of the CP-producing interaction between the task and neural net than is possible in human studies, it may be possible to model a task resembling natural speech perception and investigate what factors enable CP in such a task.

REFERENCES

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84*, 413–451. doi:10.1037/0033-295X.84.5.413

Casey, M. C., & Sowden, P. T. (2012). Modeling learned categorical perception in human vision. *Neural Networks, 33*, 114–126. doi:10.1016/j.neunet.2012.05.001

Damper, R., & Harnad, S. (2000). Neural network models of categorical perception. *Attention, Perception, & Psychophysics, 62*, 843–867. doi:10.3758/BF03206927

Gerrits, P. A. M. (2001). *The categorisation of speech sounds by adults and children.* LOT Utrecht, The Netherlands.

Gerrits, E., & Schouten, M. (2004). Categorical perception depends on the discrimination task. *Attention, Perception, & Psychophysics, 66*, 363–376. doi:10.3758/BF03194885

Harnad, S., Hanson, S. J., & Lubin, J. (1991). Categorical perception and the evolution of supervised learning in neural nets. In D. W. Powers & L. Reeker (Eds.), *Working papers of the AAAI spring symposium on machine learning of natural language and ontology* (pp. 65–74).

Harnad, S., Hanson, S. J., & Lubin, J. (1995). Learned categorical perception in neural nets: implications for symbol grounding. In V. Honavar & L. Uhr (Eds.), *Symbol processors and connectionist network models in artificial intelligence and cognitive modelling: steps toward principled integration* (pp. 191–206). London: Academic Press.

Levin, D., & Beale, J. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Attention, Perception, & Psychophysics, 62*, 386–401. doi:10.3758/BF03205558

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368. doi:10.1037/h0044417

Locke, S., & Kellar, L. (1973). Categorical perception in a non-linguistic mode. *Cortex, 9*, 355–369.

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication, 2*, 15–35. doi:10.1016/0167-6393(83)90061-4

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2*, 131–157. doi: 10.1016/0010-0285(71)90006-5

Pastore, R., Li, X.-F., & Layer, J. (1990). Categorical perception of nonspeech chirps and bleats. *Attention, Perception, & Psychophysics, 48*, 151–156. doi:10.3758/BF03207082

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition, 3*, 7–18. doi:10.3758/BF03198202

Pont, M. J., & Damper, R. I. (1991). A computational model of afferent neural activity from the cochlea to the dorsal acoustic stria. *The Journal of the Acoustical Society of America, 89*, 1213–1228. doi:10.1121/1.400652

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–535.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9*, 75–112. doi:10.1016/S0364-0213(85)80010-0

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*, 71–80. doi:10.1016/S0167-6393(02)00094-8

Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences, 10*, 538–545. doi:10.1016/j.tics.2006.10.007

Treisman, M., Faulkner, A., Naish, P. L. N., & Rosner, B. S. (1995). Voice-onset time and tone-onset time: the role of criterion-setting mechanisms in categorical perception. *The Quarterly Journal of Experimental Psychology Section A, 48*, 334–366. doi:10.1080/14640749508401394

Van Hessen, A. J., & Schouten, M. E. (1999). Categorical perception as a function of stimulus quality. *Phonetica, 56*, 56–72. doi:28441

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*, 349–355. doi:10.2307/1129249