Categorical Perception: Important Phenomenon or Lasting Myth?

Dominic W. Massaro

Perceptual Science Laboratory (http://mambo.ucsc.edu/psl/pslfan.html)
University of California
Santa Cruz, CA 95064
massaro@fuzzy.ucsc.edu

ABSTRACT

Categorical perception, or the perceived equality of instances within a phoneme category, has been a central concept in the experimental and theoretical investigation of speech perception. It can be found as fact in most introductory textbooks in perception, cognition, linguistics and cognitive science. This paper analyzes the reasons for the persistent endurance of this concept. A variety of empirical and theoretical research findings are described in order to inform and hopefully to provide a more critical look at this pervasive concept. Given the demise of categorical perception, it is necessary to shift our theoretical focus to how multiple sources of continuous information are processed to support the perception of spoken language.

1. SETTING THE STAGE

I am a glutton for punishment. What sane person would try to denigrate a sacred belief held by so many respected researchers? I should heed Leo Tolstoy's insight into the persistence of incorrect ideas throughout history. "I know that most men, including those at ease with problems of the greatest complexity, can seldom accept even the simplest and most obvious truth if it be such to oblige them to admit the falsity of conclusions which they have delighted in explaining to colleagues, which they have proudly taught to others, and which they have woven, thread by thread, into the fabric of their lives." Our human nature also biases us to seek explanations that provide easy solutions rather than those that are necessarily consistent with the observed facts.

The sacred idea I am criticizing is categorical perception (CP), or the perceived equality of instances within a category. The CP of phonemes has been a central concept in the experimental and theoretical investigation of speech perception and has also spilled over into other domains such as face processing [1]. CP was operationalized in terms of discrimination performance being limited by identification performance. Over 40 years ago, researchers at Haskins Laboratories [2] used synthetic speech to generate a series of 14 consonant-vowel syllables going from /be/ to /de/ to /ge/ (/e/ as in gate). The onset frequency of the second formant transition of the initial consonant was changed in equal steps to produce the continuum. In the identification task, observers identified random presentations of the sounds as /b/, /d/, or /g/. The discrimination task used the ABX paradigm. Three stimuli were presented in the order ABX; A and B always differed and X was identical to either A or B. Observers were instructed to indicate whether X was equal to A or B. This judgment was supposedly based on auditory discrimination in that observers were instructed to use whatever auditory differences they could perceive.

The experiment was designed to test the hypothesis that listeners can discriminate the syllables only to the extent that they can recognize them as different phoneme categories. The CP hypothesis was quantified in order to predict discrimination performance from the identification judgments. The authors concluded that discrimination performance was fairly well predicted by identification. This rough correspondence between identification and discrimination has provided the major source of support for CP.

2. RESEARCH STRATEGY

Research in the study of CP has remained oblivious to the valuable scientific strategies of Karl Popper [3] and John Platt [4]. To provide a proper assessment of any theory, it is necessary to determine how closely the predicted performance matches what is observed and to compare the accuracy of this prediction with other the predictions of other theories. When this strategy is followed, one immediately notices just how poorly the categorical describes the results. The problem is that observed discrimination is almost always better than that predicted by identification. For some reason, however, this discrepancy has never been a deterrent for advocates of CP nor has it been a central result for any alternative view.

Another important barrier to advocates of CP is to insure that discrimination performance is not simply mediated by implicit identification. That is, we always face the possibility that participants are making their discrimination judgments on the basis of identification rather than on their auditory discrimination. Some tasks are more conducive to such a mediated identification process than others.

Investigators of CP have been wedded to a verification strategy in terms of simply looking for results that agree with their initial intuition. With hindsight, we can observe that a quantitative test of any alternative theory would have described the results equally well. The inattention to alternative theories was particularly detrimental to advocates of CP, because they failed to learn that theories such as TSD and FLMP (grounded in continuous information) also predict an orderly relationship between identification and discrimination. Thus, good scientific

practice would have nipped CP in the bud, and scientists would have had more time to spend with their families.

Not only did scientists see CP where it didn't exist, they also refused to accept negative findings. Consider a very simple demonstration by Barclay [5]. Using a three category continuum from B to D to G, he limited his observers to the response alternatives B and G. If the perception of D was indeed categorical, the responses to instances in this category should have been random. However, they were instead systematically related to their stimulus properties. This study was only one of many falsifications, but the belief in CP did not diminish.

3. OUTCOME VERSUS PROCESS

Even though the concept of CP has been controversial almost from its beginning, it has survived this controversy and can be found as fact in most introductory textbooks in perception, cognition, linguistics, and cognitive science. I believe that one of the main contributions to this lasting influence is that students of speech perception have equated the necessary outcome of speech perception with the processes that led up to that outcome. No one denies the fact that speech perception requires categorical decisions. When a mother points to a toy and asks her daughter to bring the ball, the daughter must decide between the ball and a nearby doll. There must be no ambiguity in her response.

On the other hand, there is no reason why the child has only categorical information about the message. Within the framework of the fuzzy logical model of perception (FLMP), we have argued that speech perception is influenced by multiple sources of information [6]. The use of multiple sources of information in perception necessitates the fact that the sources information are continuous rather than categorical. If a source of information (such as an acoustic feature of speech) is perceived categorically, it is difficult to conceptualize how that feature would be integrated productively with other sources of information such as visible speech or linguistic context. Sentential context, for example, would either agree or disagree with the categorization of the speech input. If the sentence context agrees with the speech input, it can provide no additional information. If the sentence context disagrees with the categorization of the speech input, however, the perceiver is faced with a conflicting situation in which the context and acoustic input are inconsistent with one another. It is important to note that these logical arguments are not the only reasons that we reject CP.

4. MIMICKING CP

It is easy to create a situation to produce CP even though these results are not representative of speech perception more generally. Although speech perception is continuous, there may be a few speech contrasts that qualify for a weak form of CP. This weak form of CP would be reflected in somewhat better discrimination between instances from different categories than between instances within the same category. As an example, consider an auditory /ba/ to /da/ continuum. The F2 and F3 transitions are varied in linear steps between the two endpoints of the continuum. The syllable /ba/ is characterized by rising transitions and /da/ by falling transitions. Subjects might discriminate a rising from a falling transition more easily than discriminating two rising or two falling transitions even though the frequency difference is identical in the two cases. Direction of pitch change is more discriminable than the exact magnitude of change. This weak form of CP would be due to a fundamental characteristic of auditory processing and would not be a result of having speech categories. Thus similar results would be found in humans, chinchillas, and monkeys and for nonspeech analogs. However it is important to note that discrimination between instances within a category is still possible. In this regard, the putative CP found with non-speech is no more convincing then the results found in the speech domain. Although this weak form of CP might exist for a few distinctions, the majority of speech distinctions do not have this property [7].

5. CATEGORICAL PARTITION

I cannot understand why categorization behavior continues to be interpreted as evidence for CP. It is only natural that continuous perception should lead to sharp category boundaries along a stimulus continuum [8]. Given a stimulus continuum that is perceived continuously, we can define **i** as an index of the degree to which the information represents a particular category **I**. An optimal decision rule for making a discrete judgment would set a criterion value and classify the pattern as the category **I** for any value greater than this value. Given this decision rule, the probability of an **I** categorization would appear as a step-like function across the stimulus continuum. That is, with a fixed criterion value and no variability, the decision operation changes the continuous function given by the perceptual operation into a step function. Although based on continuous perception, this function is identical to the idealized form of CP in a speech identification task It follows that a step function for identification is *not* evidence for CP because it can also occur given continuous information. Categorical decisions made on the basis of continuous information can produce identification functions with sharp boundaries, and therefore cannot be taken to represent CP. Strictly speaking, of course, CP was considered present only if discrimination behavior did not exceed that predicted from categorization. However, one should not have been impressed that discrimination performance did not exceed that predicted by categorization if the discrimination task resembled something more akin to categorization than discrimination.

6. CONTINUOUS MEASURES

We have accumulated, as have other investigators, a variety of sources of evidence against the concept of categorical speech perception. One approach to the question of categorical speech perception is the use of continuous rather than discrete perceptual judgments. Relative to discrete judgments, continuous judgments provide a more direct

measure of the listener's perceptual experience. For example, scientists have found that a binary response proved insensitive to the manipulation of an independent variable whereas confidence ratings revealed significant effects of this variable. In these tasks, subjects were asked to rate the degree to which they felt that the speech stimulus represented one alternative or the other, rather than simply indicating which alternative was presented. Categorical and continuous models of speech perception can be formalized and evaluated against the distribution of repeated rating responses to each test stimulus along a synthetic speech continuum [9]. Categorical and continuous models of speech perception make different predictions about the distribution of repeated rating judgments to a given stimulus along some speech continuum. The results of both synthetic auditory and synthetic visual speech studies provide conclusive evidence that there is continuous information available in speech perception. In agreement with these observations, bimodal speech is also perceived continuously rather than categorically [8].

7. REINVENTING CP

Categorical perception continues to be reinvented in new disguises. Researchers could short-circuit the staleness of CP by relabeling it (perhaps categoricaly).

7.1 Perceptual Equivalence

There was a short bout of enthusiam with so-called perceptual equivalence. Consider an experiment in which the silent closure duration and the vocalic formant transition onsets were independently varied to cue the distinction between the words *slit* and *split* [10]. Silence between the noise of the initial /s/ and the onset of the vocalic portion of the word is a cue for /p/. Rising formant contours at the onset of the vocalic portion is also a cue for /p/. Conversely, little or no silence and flat formants are cues for *slit* rather than *split*. Different values of silence and formant contours can be chosen to produce different stimuli that are identified equivalently in a labeling task. A speech stimulus with a silence of 72 ms and with rising formant contours was identified as *split* about 82 percent of the time. The same identification proportion was found for a speech stimulus with a silence of 104 ms and with flat formant contours. According to proponents of perceptual equivalence, these two speech stimuli are perceptually equivalent and difficult to discriminate from one another. However, we know that these two items are easy to discriminate from one another.

7.2 Perceptual Magnets

More recently, the perceptual-magnet effect (PME) has had a tremendous impact on the field, and has generated a great deal of research [8]. The critical idea is that the discriminability of a speech segment is inversely related to its category goodness. Ideal instances of a category are supposedly very difficult to distinguish from one another relative to poor instances of the category. If we understand that poor instances of one category will often tend to be at the boundary between two categories, then the PME is more or less a reformulation of prototypical CP. That is, discrimination is predicted to be more accurate between categories than within categories. In demonstrating its viability, the PME faces the same barriers that have been difficult to eliminate in CP research. In standard CP research, it is necessary to show how discrimination is directly predicted by identification performance. In the PME framework, it is also necessary to show how discrimination is directly predicted by a measure of category goodness. We can expect category goodness to be related to identification performance. Good category instances will tend to be identified equivalently, whereas poor instances will likely be identified as instances of different categories. Lotto et al. [11] observed that category goodness ratings are context sensitive in the same manner than identification is. To control for this influence, they obtained the category goodness ratings in the same presentation context as the identification judgments. They found that there was a direct correspondence between the goodness ratings and the identification judgments. This correspondence is consistent with theories of continuous perception.

8. IMPLICATIONS FOR INQUIRY

Notwithstanding the three decades of misinterpreting the relationship between identification and discrimination of auditory speech, we must conclude that it is perceived continuously and not categorically [8]. Our research reveals that visible and bimodal speech are also perceived continuously. This observation pulls the carpet from under current views of language acquisition that attribute to the infant and child discrete speech categories [13,14]. Most importantly, the case for the specialization of speech is weakened considerably because of the central role that the assumption of CP has played [15]. Finally, several neural network theories such as single-layer perceptrons, recurrent network models, and interactive activation have been developed to predict CP: its nonexistence poses great problems for these models.

9. THE FUZZY LOGICAL MODEL

Given the demise of CP, we are now faced with a more challenging situation in that we must address how multiple continuous sources of information are evaluated and integrated to achieve a percept with continuous information. Keith Kluender, a talented and productive researcher, who has studied speech perception from almost all possible perspectives, summarized his understanding in the following manner [16]. There are no auditory discontinuities in speech; each distinction has multiple stimulus attributes; and experience is critical. These conclusions are the bedrock of the FLMP.