JASA ARTICLE

The myth of categorical perception[a)]

Bob McMurray[b)]
Department of Psychological and Brain Sciences, University of Iowa, Iowa City, Iowa 52242, USA

ABSTRACT:
Categorical perception (CP) is likely the single finding from speech perception with the biggest impact on cognitive science. However, within speech perception, it is widely known to be an artifact of task demands. CP is empirically defined as a relationship between phoneme identification and discrimination. As discrimination tasks do not appear to require categorization, this was thought to support the claim that listeners perceive speech solely in terms of linguistic categories. However, 50 years of work using discrimination tasks, priming, the visual world paradigm, and event related potentials has rejected the strongest forms of CP and provided little strong evidence for any form of it. This paper reviews the origins and impact of this scientific meme and the work challenging it. It discusses work showing that the encoding of auditory input is largely continuous, not categorical, and describes the modern theoretical synthesis in which listeners preserve fine-grained detail to enable more flexible processing. This synthesis is fundamentally inconsistent with CP. This leads to a different understanding of how to use and interpret the most basic paradigms in speech perception—phoneme identification along a continuum—and has implications for understanding language and hearing disorders, development, and multilingualism. © *2022 Acoustical Society of America.*
https://doi.org/10.1121/10.0016614

(Received 21 February 2022; revised 26 November 2022; accepted 6 December 2022; published online 29 December 2022)
[Editor: Matthew B. Winn] Pages: 3819–3842

I. INTRODUCTION

Throughout the history of speech perception research, categorical perception (CP; Liberman *et al.*, 1957) has been a breakout finding. It is one of only a few that have made a lasting impact outside of the field (Goldstone and Hendrickson, 2010; Harnad, 1987), and its influence continues to grow (Fig. 1). CP is an empirical phenomenon and a theoretical claim about perceptual encoding. Empirically, CP is observed when discrimination (the ability to distinguish two stimuli) is affected by the presence of categories. Theoretically, it claims that perception is "warped" by the presence of categories and "analog or continuous inputs are transformed into quasi-digital, quasi-symbolic encodings" (paraphrased from Goldstone and Hendrickson, 2010).

CP is wrong at both levels. Researchers actively working on speech categorization have known this for many years (e.g., Massaro and Cohen, 1983; Pisoni and Lazarus, 1974). Yet, long after the seminal work ruling out CP, it continues to live on as a sticky scientific meme. It influences the way speech perception is conceived in work on language development (Maurer and Werker, 2014), communication disorders (Serniclaes *et al.*, 2004), and neuroscience (Chang *et al.*, 2010) as it shapes ideas about what an ideal listener should be aiming for. CP continues to appear in other areas of perception (Beale and Keil, 1995; Franklin *et al.*, 2008; Freedman *et al.*, 2001) and other species (Green *et al.*, 2020; May *et al.*, 1989). It has left us with problematic tasks still used today for many purposes.

Despite the substantial work countering the claims of CP, few (if any) accessible reviews explain why it is wrong or wrestle with the consequences of the failure of CP for theories of speech perception and the methods used to study it. This may be why CP still appears unchallenged in many textbooks and why so many subfields of cognitive science do not appear to be aware of its demise. Thus, this review seeks to directly lay out the evidence against CP and for alternative conceptualizations of speech categorization. This review is not exhaustive. It focuses on work that directly examines categorization in its "distilled" form (laboratory studies of monolingual typical adults). I lack the expertise (or space) to offer more than a superficial discussion of other relevant fields such as sociolinguistics, speech pathology, development, second language (L2) learning. This is not a claim that the more distilled work emphasized here captures speech perception in general nor that it offers greater insight than these other fields. Rather, I focus on this work because this is from where the large body of results that directly challenge CP has emerged. It is this field where the critical breakdown in scientific story telling began.

This review is more blunt than typical. The widespread view among researchers working on speech categorization is that CP is finished, and there is no need to continue studying it (Crowder, 1989; Schouten *et al.*, 2003). However, the persistence of CP suggests the need for a more direct treatment. *This review does not argue that people do not*

[a)] This paper is part of a special issue on Reconsidering Classic Ideas in Speech Communication.

[b)] Also at: Department of Linguistics and Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242, USA and Haskins Laboratories, New Haven, CT 06511, USA. Electronic mail: Bob-mcmurray@uiowa.edu

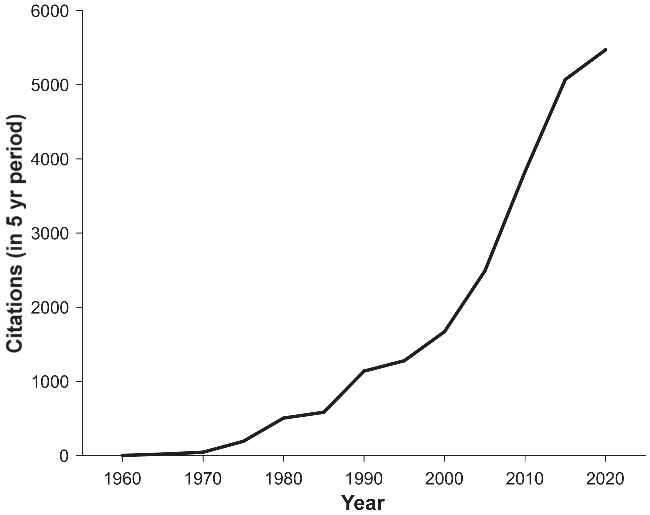


FIG. 1. The number of new citations in consecutive five-year periods starting in 1955. The search was conducted on Google/Scholar (Mountain View, CA) with the term "categorical perception."

categorize *speech*. The evidence for categorization is strong. Rather, CP is a deeper claim about perceptual representations below the level of categories, and it is those claims that are refuted here. The broader claim is that CP privileges the goal of identifying a single phonological category from a single segment over other crucial goals, both phonological (recognizing multiple phonemes in parallel) and non-phonological (word recognition and social perception). Thus, it stands as a barrier to a richer and more comprehensive understanding of speech perception.

II. CP

A. Empirical definition of CP

CP is assessed in experiments in which listeners hear tokens from a continuum of speech sounds (e.g., spanning /b/ to /p/ in small steps). CP requires tasks that assess labeling (is each token a /b/ or a /p/?) and discrimination (are two tokens the same or different?). CP is observed when three conditions are met. First, listeners must show a sharp labeling function [Fig. 2(A), dashed line]: as the speech continuum advances, there is a sudden shift in categorization. Second, listeners should be poor at discriminating tokens from the same category (solid line), and, third, they should be good at discriminating tokens that –cross the boundary.

The first criterion is difficult to assess. The psychophysical function cannot be linear. A listener cannot choose any category more than 100%. Thus, the function must reach asymptotes at either end, and any apparent linearity is an artifact of where the end points are set. The mere presence of a sigmoidal curve, in fact, means little other than that subjects had to make a forced choice decision of tokens on a continuum. It can never be evidence for CP. The steepness of the function is similarly problematic. Although the strongest form of CP claims a perfect step function (a transition of 100% between the smallest possible distance), this has never truly been tested (e.g., with adaptive approaches that use increasingly smaller steps). Instead, researchers generally informally evaluate the slope and declare it steep. Yet, slope is, in part, related to the step size and the range of the continuum (Rosen, 1979). By selectively testing a narrow range (or small step sizes), a researcher could create sharper or shallower transitions. Moreover, across speech cues, continuum steps are incommensurate [how does a 5 msec voice onset time (VOT) difference for a voicing continuum compare to a 40 Hz first formant (F1) frequency difference for a vowel continuum], and there have been few attempts to standardize them (though, see the supplementary material S1[1]). Thus, the first criterion cannot be rigorously evaluated.

The second and third criteria are often treated similarly. Some have caricatured CP as requiring *no above chance within-category discrimination*. However, even the original demonstration of CP (Liberman *et al.*, 1957) did not observe this (finding robust discrimination within the /g/ category). Nonetheless, this characterization is not baseless.

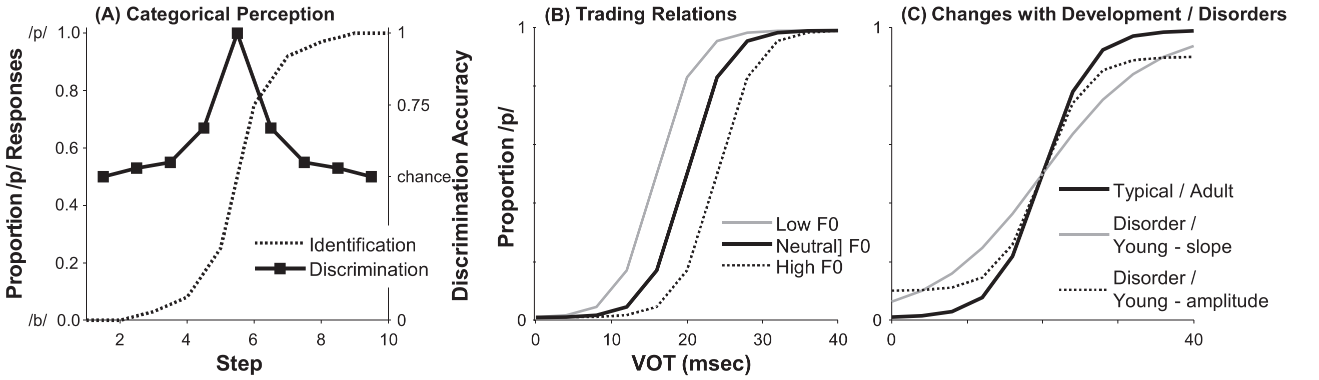


FIG. 2. (A) Canonical CP results profile. On the *x* axis, step is step from a speech continuum, where step 1 might be a prototypical /b/ and step 10 may be a prototypical /p/. Identification (dashed line, left axis) is coded as the proportion of responses that match one of the end points. Here, zero is /b/ and 1.0 is /p/, such that 0% responding indicates the participant consistently heard /b/. Discrimination is assessed between neighboring points (e.g., between steps 1 and 2, 2, and 3, and so forth). The peak indicates that discrimination is better between steps 5 and 6 (spanning the boundary in the identification results) than between steps 1 and 2 (both /b/). [(B), (C)] Schematic results of typical effects on identification alone. (B) Trading relations paradigms show how cues interact, for example, the separate influence of VOT (*x* axis) and F0 (separate curves) on voicing categorization. (C) Other paradigms focus on the slope, for example, comparing voicing categorization in younger and older children or in disordered and typical children.

Indeed, Liberman *et al.* (1961) writes in the abstract to the second CP paper: "…perception … is essentially categorical in that *S can hear no differences among the stimuli beyond those that are revealed by the phoneme labels*." Nonetheless, this high bar is not conventionally required for CP, and many collapse the second and third criteria to the *relative* difference (discrimination is *better* for tokens spanning the boundary than those within a category).

A more rigorous approach is to directly predict discrimination from labeling responses (Gerrits and Schouten, 2004; Liberman *et al.*, 1957; Schouten *et al.*, 2003). Such models start from the strong assumption that listeners perceive speech solely in terms of labels. Under this assumption, discrimination should be at chance if two tokens are labeled the same 100% of the time and should increase if labeling of those tokens differs. The predicted discrimination function is generated from each subject's labeling and compared to their obtained discrimination. To the extent that they match, CP is claimed. To the extent that discrimination exceeds predictions, some within-category sensitivity must be present. Perfect predictability need not be found to claim CP (identification and discrimination need not be fully isomorphic).

Using variants of this definition, CP has been shown across a variety of contrasts, including place of articulation (Liberman *et al.*, 1957) and voicing (Liberman *et al.*, 1961) in stop consonants, vowels (Van Hessen and Schouten, 1999; (though, see Fry *et al.*, 1962), liquids like l/r (Miyawaki *et al.*, 1975), and the fricative/affricate contrast (Cutting and Rosner, 1974).

This definition of CP has not been consistently applied. Often the term CP is used to refer to a steep labeling function alone; others use CP to refer to any experiment that asks subjects to label tokens from a continuum even with no discrimination measure. *This is wrong. We must end this rhetoric.* CP invokes a specific theoretical view of speech (now disproven). By invoking CP as a sort of shorthand for this approach, researchers are (perhaps unintentionally) making stronger claims about perception than they had intended. Even as a purely methodological term, these studies are simply measuring categorization—not CP.

In other cases, CP is sometimes claimed on the basis of discrimination data alone in the absence of a labeling task. This is common in the infant literature as most measures only permit a measure of discrimination (McMurray, 2022, for a review); however, it also applies to neuroimaging techniques that use approaches such as the mismatch negativity (MMN) in an electroencephalogram (EEG) or representational similarity analyses to construct a neural version of discrimination but without any corresponding neural identification. Such practices risk affirming the consequent: we know a particular profile of discrimination is consistent with CP; CP requires categorization; therefore, when that profile is observed, we can assume categorization. However, as I describe, the evidence against CP is now substantial. If we remove CP from the derivation chain, it is not clear what can be concluded from discrimination alone.

B. Theoretical implications of CP

CP also makes theoretical claims about speech. It starts from a model of speech recognition with two levels of analysis (Fig. 3). In the first analysis, input is mapped onto some form of auditory encoding analogous to acoustic cues such as the spectral peak of a fricative, formant frequencies of a vowel, or the VOT of a stop. This does make strong claims for the validity of any specific cue, and most likely this level reflects multiple. Critically, this level of representation is continuous: at some level, it preserves the gradient nature of the input in a way that reflects acoustic similarity. In the second level of analysis, cues are carved into categories like phonemes, features, or words. In the framing of CP, *perception* refers to the auditory/cue encoding, and categorization is a later cognitive process operating on perception. This contrasts from more modern (looser) uses of these terms in which perception might include a wide range of processes, including categorization.

Critically, discrimination tasks do not require labeling. For example, in the *ABX* task used by Liberman *et al.* (1957), listeners hear two distinct sounds (*A* and *B*) followed by a third (*X*) that matches *A* or *B*. They indicate whether the third sound matched *A* or *B*. For example, if they heard "ka ga ga," they should respond *B*. In principle, a listener could perform this task without labels; if they heard ga ga ga (where the underline indicates a lower pitch), they could identify *B* as the match. Thus, discrimination tasks were presumed to tap auditory encoding. Under this assumption, CP makes a bold claim: auditory encoding is altered or warped by perception. Tokens from distinct categories are heard as more distant in perceptual space, and tokens from the same

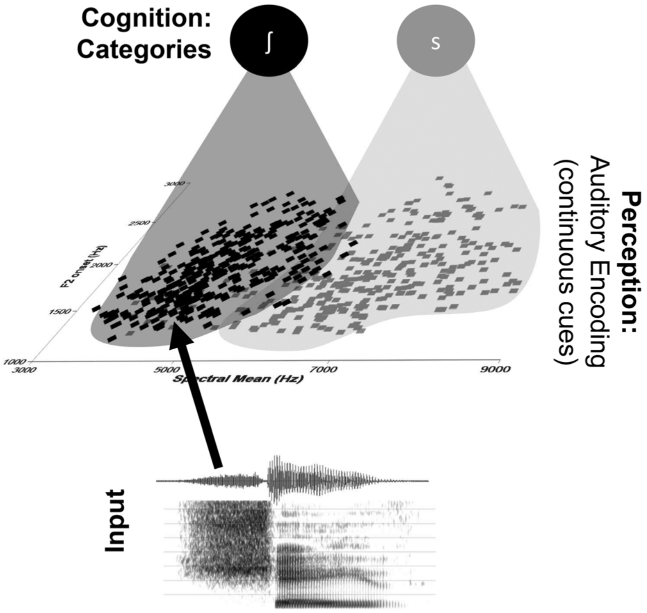


FIG. 3. A rough conceptual model of the speech chain assumed by CP. Speech input is first mapped to a continuous auditory encoding that represents something like continuous cue values. This space is then carved into categories.

category are closer. This violates Weber's law (perceptual distance is a monotonic function of stimulus distance) as perception is warped toward a more discrete representation.

The perceptual warping reflected by CP was a view that was well suited for its moment in the history of cognitive science. At the time of CP, work at Haskins Laboratories (New Haven, CT), the Massachusetts Institute of Technology (Cambridge, MA), and Brown University (Providence, RI) was defining *the problem of lack of invariance* (Blumstein and Stevens, 1979; Delattre *et al.*, 1955): the idea that variability due to talker differences, speaking rate, and coarticulation makes it such that there is no one-to-one mapping between an acoustic form and a phoneme. This is illustrated in Fig. 4(A), which shows measurements of /s/ and /ʃ/ across multiple vowels and talkers from McMurray and Jongman (2011). Note the overlap between categories driven by talker and coarticulation from the neighboring vowel. Figure 4(B) shows these same measurements transformed by a simple version of CP, which enhances between-category distance and reduces within-category variation. By this analogy, CP appeared to make the problem of invariance less challenging. It does not solve it—in fact, the tokens on the wrong side of the boundary are now misclassified in more extreme ways. However, at the time it was discovered, the empirical evidence for CP implied that listeners may be equipped with mechanisms specialized for speech to solve the problem of lack of invariance (by yet unknown means).

Second, when CP was discovered, the cognitive revolution was under way. This movement argued for a view of cognition built on symbolic operations. CP supported this view by arguing for the existence of mechanisms that rapidly transduce the continuous signals into categories (symbols). CP later became linked to other key ideas in the cognitive revolution: for example, it may be innate and specific to speech (Liberman *et al.*, 1967). Supporting this special mode of speech perception, there were early demonstrations that CP could be observed with speech but not nonspeech sounds that capture similar acoustic relationships (Liberman *et al.*, 1961; see Cutting and Rosner, 1974, for a review).

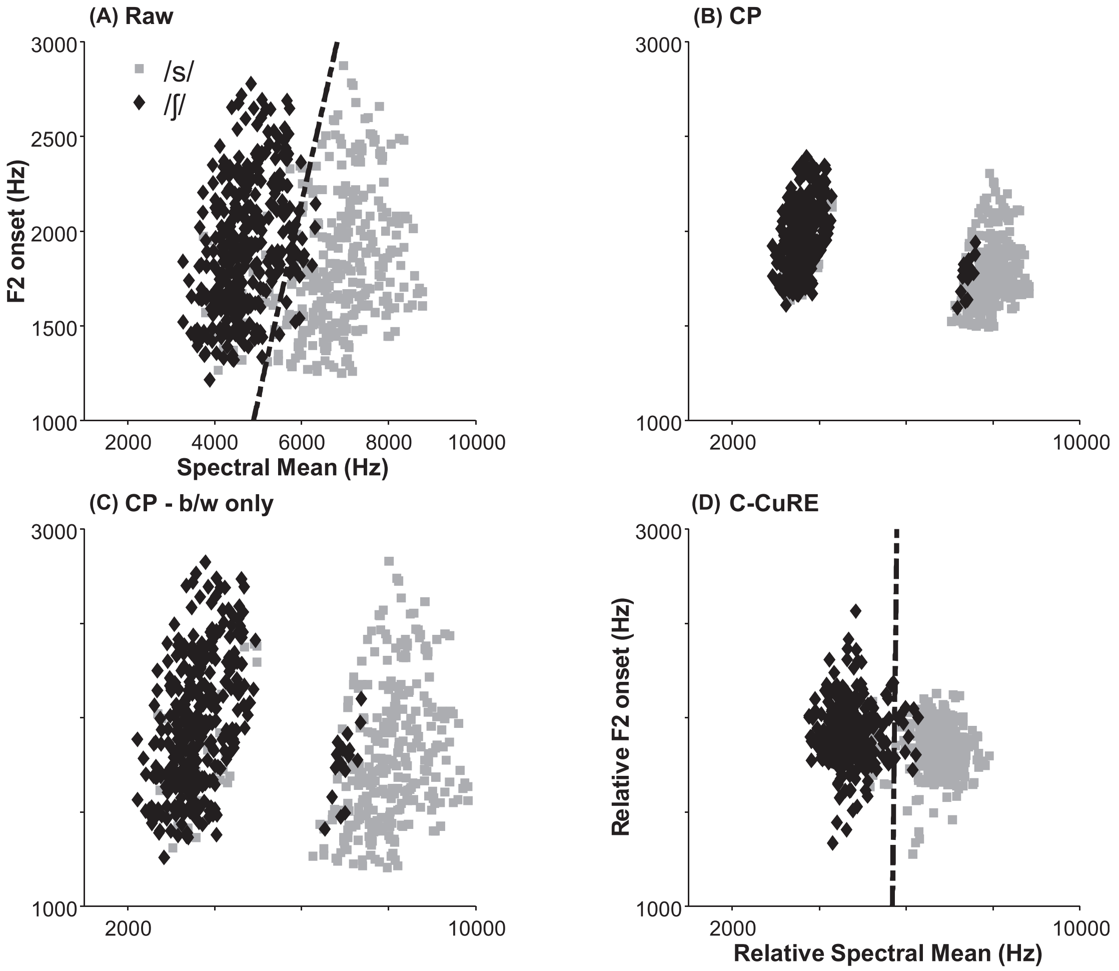


FIG. 4. (A) Spectral mean and F2 onset frequencies for /s/ and /ʃ/ from McMurray and Jongman (2011). The line represents a linear discriminant analysis that can separate the categories at about 92%. (B) A simulation of CP generated by assuming the linear boundary in (A), and then doubling the distance between tokens on each side of the boundary and halving the variation within each side of the boundary; (C) a simulation of CP generated by only increasing the distance without adjusting within-category variation; and (D) the same data after compensating for talker and vowel differences using the regression approach of C-CuRE (McMurray and Jongman, 2011) are depicted; the discriminant analysis achieves 96.6% accuracy.

Whether or not CP is innate or special to speech, the broader theoretical claim has persisted: perception of continuous acoustic representations is warped by language to support categorization. Several mechanisms have been posited to account for this.

One early account is that CP derives from natural *discontinuities in the auditory system* (e.g., discontinuities in perceiving any time difference near 20 msec, which is the typical VOT boundary); this was supported by work showing CP for speech sounds in nonhuman animals (Kuhl and Miller, 1975) and with nonspeech analogues that capture similar acoustic distinctions as speech (e.g., tone-onset-time as an analogue of VOT; Cutting and Rosner, 1974; Pisoni, 1977). Auditory discontinuities were also supported by later electrophysiological work using direct recordings from auditory cortex (Steinschneider *et al.*, 1999). This challenges the idea that CP is a marker of speech-specific processing. Under this view, languages evolved to take advantage of these discontinuities.

The discontinuity approach was unlikely to fully explain the phenomenon. First, these studies focused on a handful of phonemic contrasts (largely voicing and place of articulation in stop consonants), and it is unclear if auditory discontinuities are present for other contrasts. Moreover, languages vary in where they place their boundaries, and it is unclear if discontinuities would line up crosslinguistically. Second, some empirical results with nonspeech analogues may derive from methodological confounds (cf. Rosen and Howell, 1981). Third, animal models are problematic as animals must be extensively trained and, consequently, CP could be an artifact of learning not an inherent property of the auditory system.

One compelling alternative account was that the auditory space is *warped by learning*. Overtraining on particular classes of sounds (over development) led the system to devote more representational or neural space to certain subregions of the acoustic space (Guenther and Gjaja, 1996). This still posits the same discontinuity in the auditory system, but its origin is learning. This is more parsimonious than prior views as it can apply to any phoneme contrast or boundaries used crosslinguistically. It is supported by work showing that discrimination peaks align with the language of the listener: Japanese listeners show no discrimination peaks in an /l/-/r/ continuum while English speakers do (Miyawaki *et al.*, 1975). This has been instantiated in neural models (Guenther and Gjaja, 1996) and empirically examined with magnetic resonance imaging (MRI; Guenther *et al.*, 2004), suggesting plasticity in neural "maps" such that more space is given to categories.

In a similar vein, CP may derive from an *interactive feedback loop* (Anderson *et al.*, 1977; see also Lupyan, 2012). As the system settles on one choice, activation at the category level feeds back to perceptual encoding, aligning the perceptual representation with the category decision. This may be an indirect effect of learning (because categories are still learned), but learning does not alter the structure of the auditory representation; rather, it is functionally altered in real time as perception and cognition align.

Whatever the theory, the strong assumption of CP is that the perceptual space (defined as the pre-categorical, auditory encoding) is at least partially discontinuous. The depth of this assumption is perhaps illustrated by the TRACE model (McClelland and Elman, 1986). This model was built on fundamentally different assumptions about cognition from the cognitive revolution, rejecting the need for discrete symbols and the specialness of speech. Yet, the authors still felt the need to extensively explain CP as a product of competition (McClelland and Elman, 1986, pp. 43–50). This has left a consensus that whatever the mechanism, auditory encoding does not faithfully reflect the input and is warped in a way to rapidly transduce auditory input to something quasi-discrete.

C. Beyond speech

This formulation has extended beyond speech. CP has been shown in complex auditory domains such as musical categories (Howard *et al.*, 1992). It has been invoked in vision, including for lower-level cues such as color (Bornstein and Korda, 1984) and line orientation (Quinn, 2004), and higher-level stimuli such as faces (Beale and Keil, 1995), facial emotion (Hess *et al.*, 2009), and object categories (Newell and Bülthoff, 2002). It has been shown in the tactile and haptic (touch) perceptual systems (Gaißert *et al.*, 2012; Knight *et al.*, 2014). It has even been observed in students' judgements of *p*-values (Rao *et al.*, 2022), in which *perception* is hardly relevant! It has been shown in several species, often for species-relevant signals (e.g., birdsong; Lachlan and Nowicki, 2015; May *et al.*, 1989; Wyttenbach *et al.*, 1996).

This work outside of speech has enriched the debate. It would be surprising if fundamental perceptual discontinuities appeared in all of these domains. Moreover, this research has revealed unequivocally that CP is at least a partially learned phenomenon. For example, CP appears for familiar but not unfamiliar faces (Beale and Keil, 1995), and it can be invoked by learning new faces (Goldstone *et al.*, 2001). It appears in musical chords only for musically trained people (Howard *et al.*, 1992). CP in color vision is linked to the specific color terms of the participant's language (Roberson *et al.*, 2000) and developmentally linked to learning color words (Franklin *et al.*, 2008).

In these domains, CP often supports similar claims as in speech: that a given domain is handled by a system that transduces a continuous sensory space to categories (Knight *et al.*, 2014), often by specialized mechanisms. For example, inverted faces are often used to probe face-specific processing, and CP is reduced in inverted faces (McKone *et al.*, 2001; though, see Levin and Beale, 2000). Similarly, hemispheric asymmetries are often invoked to argue for a role of language, and color CP is stronger when stimuli are presented to the right visual hemifield (Franklin *et al.*, 2008), which projects most strongly to the left hemisphere. The power of the CP metaphor has led to the claim that it is foundational to cognition (Harnad, 1987).

D. But. . .

CP offers a compelling package. The work reviewed so far suggests that CP is a domain-general (and species-general) mechanism in which learning warps the perceptual space in a way that may help deal with the variability in the input.

Unfortunately, it is wrong.

The empirical evidence for CP is not strong, and the theoretical views it inspired are fundamentally inconsistent with modern theories of speech. This does not rule out categorization as an essential aspect of speech perception. Rather, what is wrong is the deeper claim about auditory/perceptual encoding. To make this argument, I address three claims. The first claim is the evidence for auditory warping (Sec. III). The second claim is the broader claim about the quasi-discrete nature of categories (Sec. IV). Although this latter point is not direct evidence against CP, it is evidence against the broader view of speech rooted in CP. Finally, I probe broader facts about speech and cognition, which suggest CP is simply not consistent with the kinds of mechanisms listeners must employ to solve known problems in speech (and beyond).

III. CP IS AN INACCURATE EMPIRICAL DESCRIPTION OF PERCEPTION

A. CP is not universal across speech sounds

From the outset, it was known that CP was not observed for all speech contrasts. Early studies with vowels did not show CP (Fry *et al.*, 1962; yet, see Van Hessen and Schouten, 1999) with no difference between within- and between-category discriminations. Fricatives showed reduced CP (Healy and Repp, 1982), a smaller difference than expected. More recent work shows non-CP for Mandarin tones (Francis *et al.*, 2003). Idiosyncratic findings of CP are also seen in vision: Newell and Bülthoff (2002) report CP across only some continua.

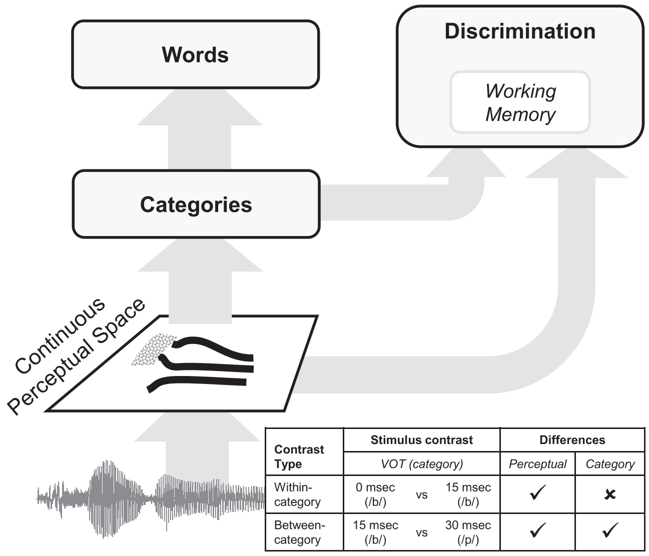
It might be easy to chalk these cases up as exceptions to a general principle of CP. This is problematic for three reasons. First, the number of speech sounds for which CP has been shown is not large. Much of the work has examined only stop consonants, and large swaths of the phonetic space have not been systematically examined: many vowel contrasts, nasals, some approximants, and some manner of articulation distinctions. Consequently, it is not clear if the contrasts showing non-CP are the exceptions or if the cases showing CP are the exceptions. Second, if CP derives from a learned warping or feedback (dominant accounts at the moment), why do these not operate for all sounds? There is no clear theory that explains how the general principles of CP interact with specific phonetic cues to yield such idiosyncratic effects. Third, these findings are inconsistent with the theoretical view that CP reflects adaptations that help cope with the problem of lack of invariance. Vowels, fricatives, and tones show some of the strongest contextual dependencies due to talker and coarticulation—these are the contrasts for which lack of invariance is a more serious problem. If CP is supposed to be part of the solution to the lack of invariance, why is it not operative for these contrasts?

B. Discrimination is not categorical

A more direct challenge to CP comes from work explicitly assessing discrimination tasks. Classic work used the *ABX* task. Ostensibly, this task does not require labeling. However, it has a high memory load. Listeners must encode the *A* and *B* stimulus and retain them to compare with the *X* stimulus. A phonological code may be a more efficient and durable form of encoding than an auditory one—if the auditory code fades, all that may be left is the categorical code. Under this view, the discrimination peaks of CP do not reflect CP but, rather, categorical *memory* or categorical *judgements* (Pisoni, 1973).

In fact, detailed analyses of the discrimination tasks used to assess CP (Macmillan *et al.*, 1977; Pollack and Pisoni, 1971) led to the conclusion that the *ABX* and *AX* (same/different) tasks were perhaps the least suited for assessing pre-categorical auditory encoding. Both are memory intensive and have unclear response criteria (e.g., a subject could choose to respond at a phonemic or auditory level). This led to discrimination tasks that are less memory intensive or biased, such as the oddball, 4IAX, and 2IFC tasks (see the supplementary material S2 for details[1]). These tasks show substantial within-category discrimination and no discrimination peak at the boundary (Carney *et al.*, 1977; Pisoni and Lazarus, 1974). Most impressively, Gerrits and Schouten (2004) and Schouten *et al.* (2003) first established nearly perfect CP with a standard *ABX* task. When they then switched to the 4IAX/2IFC task, there was no evidence for CP—discrimination was not predicted by identification. Thus, even playing by the rules of CP, there is no evidence for a warping of discrimination when we properly understand the demands of discrimination tasks.

Finally, Pisoni offers an insightful analysis that undercuts the premise of using *any* discrimination task to infer CP (Pisoni, 1973; Pisoni and Tash, 1974). He starts from a conceptualization of speech perception as a series of transformations from the continuous perceptual signal to categories (see the left pathway in Fig. 5). CP assumes that discrimination tasks primarily tap the continuous perceptual space. However, what if discrimination judgements involve perceptual and category information (see the right pathway in Fig. 5)? Under this assumption, a within-category contrast has essentially one source of evidence favoring discrimination (any auditory difference), whereas a between-category contrast has two sources of evidence favoring discrimination (the auditory difference and the label difference; see the table in Fig. 5). That is, even if the internal perceptual difference is equivalent for within- and between-category



Contrast Type	Stimulus contrast			Differences	
	VOT (category)			*Perceptual*	*Category*
Within-category	0 msec (/b/)	vs	15 msec (/b/)	✓	✗
Between-category	15 msec (/b/)	vs	30 msec (/p/)	✓	✓

FIG. 5. The Pisoni and Tash (1974) model of discrimination. Speech is transduced to a continuous perceptual space and then to categories and words (left branch). Discrimination, however, can use either the perceptual or category representation (right branch). Under these assumptions, between-category contrasts should almost always be more discriminable than within (embedded table), even if the perceptual encoding is continuous.

contrasts (no sensory warping), one should still observe CP. Critically, in this model, it is not necessary to posit that categorization occurs before or after perceptual encoding. As long as a perceptual and a category representation are available when the discrimination response is made, the predictions hold.

This model has a radical consequence: CP should be evident *even if auditory perception is continuous!* Any differences among discrimination tasks may derive from differences in the degree to which a given task (or stimulus) differentially emphasizes category-level or perceptual-level information when making the judgement. This is a powerful argument—largely ignored by much of the literature—suggesting a fundamental limit in the degree to which the combination of identification and discrimination can tell us anything about CP as a theory of speech (see also Massaro and Hary, 1984).

To summarize, some discrimination tasks (*ABX* and *AX*) do not accurately isolate pre-categorical perceptual encoding; and in many cases, a CP-like profile of discrimination can be predicted even in the absence of true CP. Consequently, there is little clear unambiguous positive evidence for CP from the very tasks that define it. This prompts the need for methods to better isolate auditory encoding.

C. And perception is continuous (mostly)

Massaro and Cohen (1983) used an innovative way to assess the auditory encoding more directly. Subjects used a continuous rating scale to rate how /b/- or /p/-like the stimulus was using a continuous rating task, which is now referred to as a visual analogue scale (VAS; they also tested place of articulation and vowels). When averaged across trials, such measures should show the standard sigmoidal function. However, a sigmoidal function could derive from one of two underlying models (Fig. 6), which can be revealed by examining the distribution of responses. If listeners encode speech categorically, they should not hear continuous differences between intermediate continuum steps. Rather, they should consistently choose a low rating for /b/ and a high rating for /p/ (for example) because they do not hear differences in the middle of the range. Any differences as the stimulus advances along the continuum should be driven by how likely a high or low value is chosen [Fig. 6(A)]. On every trial, listeners hear either a /b/ or /p/, but the relative likelihood of those categories varies across the continua. Alternatively, if listeners encode the continuous difference between tokens, one might expect the mean rating to advance linearly as the step increases [Fig. 6(B)]. Massaro and Cohen (1983) compared the fit of these models to individual subjects and found that the continuous model offered a better fit for most subjects. This provides strong evidence of a continuous underlying percept. If it was warped, listeners would not have access to the continuous information for their ratings (for replications, see Kapnoula *et al.*, 2021; Kapnoula and McMurray, 2021; Kapnoula *et al.*, 2017; Kong and Edwards, 2011; and see Apfelbaum *et al.*, 2022, for further discussion of this task and the limits of the 2AFC task).

More recent work used electrophysiology to assess auditory encoding more directly. Toscano *et al.* (2010) measured event related potentials (ERPs) from the scalp while listeners categorized tokens from a VOT continuum. They examined the N1, an early negative deflection of the waveform that has been linked to the first cortical processing of sound. They found a linear relationship between the VOT of the stimulus and amplitude of the N1 (Fig. 7): shorter VOTs showed more negative N1s, and N1 increased linearly with VOT. Critically, there was no effect of the subject's response and no evidence for warping near the boundary (see Sarrett *et al.*, 2020; Toscano *et al.*, 2018, for a replication; Getz and Toscano, 2021, for a review). This offers clear evidence that at the earliest stages of perception, the encoding of speech is not warped by the presence of categories.

Given these results, what can be made about neuroscience results suggesting warping? For example, direct recordings from humans' auditory cortex (Steinschneider *et al.*, 1999) show a discontinuity such that lower VOTs have a single negative deflection while longer VOTs have two negative deflections. However, for long VOTs, the two deflections linearly tracked the VOT, suggesting gradiency, and at lower VOTs, the two peaks may be close enough to smear together. Similarly, Chang *et al.* (2010) used a representational similarity analysis on recordings from the human superior temporal gyrus (STG) to demonstrate that tokens within a category were closer than those spanning a boundary (the classic definition of CP). However, that cortical region is also a locus of phonological processing

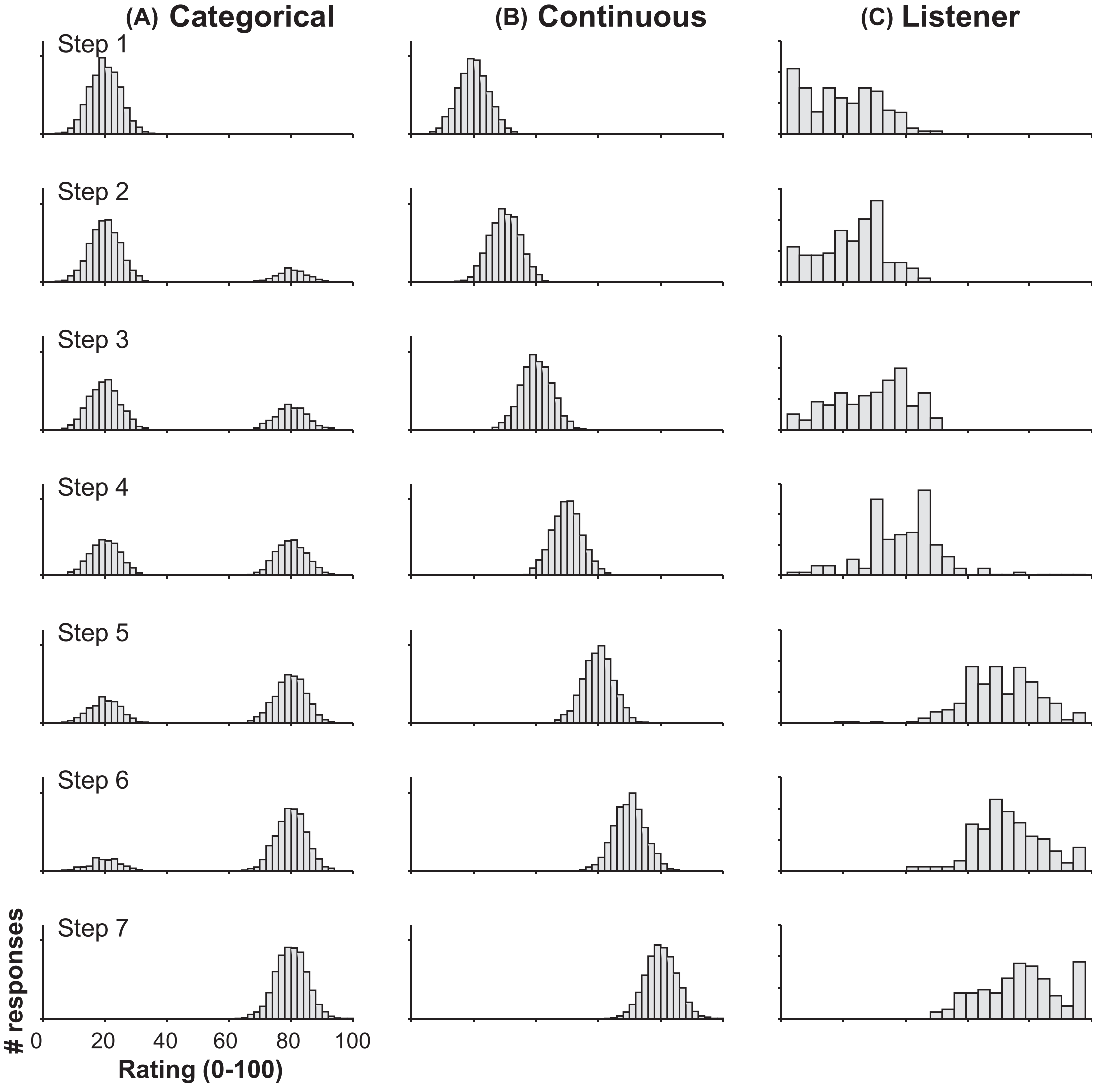


FIG. 6. Distribution of rating scale responses for individual subjects as a function of continuum step (each row) from Massaro and Cohen (1983; estimated from their Fig. 7). (A) Under a categorical model, listeners should use only the ends of the scale and only vary in how frequently they choose the /b/ or /p/ ends. (B) Under a continuous model, the mean rating should shift with step. (C) Estimated data from a typical listener is shown. Column (C) reprinted with permission from Massaro and Cohen, Speech Commun. 2, 15–35 (1983). Copyright 1983 Elsevier.

(Mesgarani *et al.*, 2014), raising the possibility that this measure does not isolate auditory encoding but simultaneously reflects category representations (consistent with Pisoni and Tash, 1974, Fig. 5). Indeed, Pasley *et al.* (2012) used similar recordings from STG to reconstruct the spectro/temporal form of zspeech. This would not have been possible with the information loss implied by CP.

Finally, an intriguing challenge to linear cue encoding comes from work by Kapnoula and McMurray (2021). They also applied the N1 paradigm of Toscano *et al.* (2010) but examined individual differences by relating the N1 response function to the degree of gradiency shown in a continuous rating (VAS) task (e.g., Fig. 4). Kapnoula and McMurray (2021) found that more categorical listeners showed a small bump in the N1 function at the category boundary while gradient listeners were linear. However, even the categorical listeners still encoded continuous detail *within each category*. Thus, the presumed acoustic space of this subset of listeners would look more like the visualization in Fig. 4(C) in which the distance between categories is expanded but the ability to represent within-category detail is unchanged. Such models—which may only apply in some listeners—are not commonly considered in the CP pantheon. However, no existing models are equipped to account for such differences at an individual level, raising the need for further investigation.

D. Meanwhile in other subdomains of speech

Outside of the community of researchers actively working on CP, CP offers a less than compelling model that is often bypassed to do useful work. For example, in sociolinguistics, within-category variation often signals identity (gender, dialect, and individuals). If auditory representations were warped to minimize such variation, how could this important function of speech be conveyed, and how could

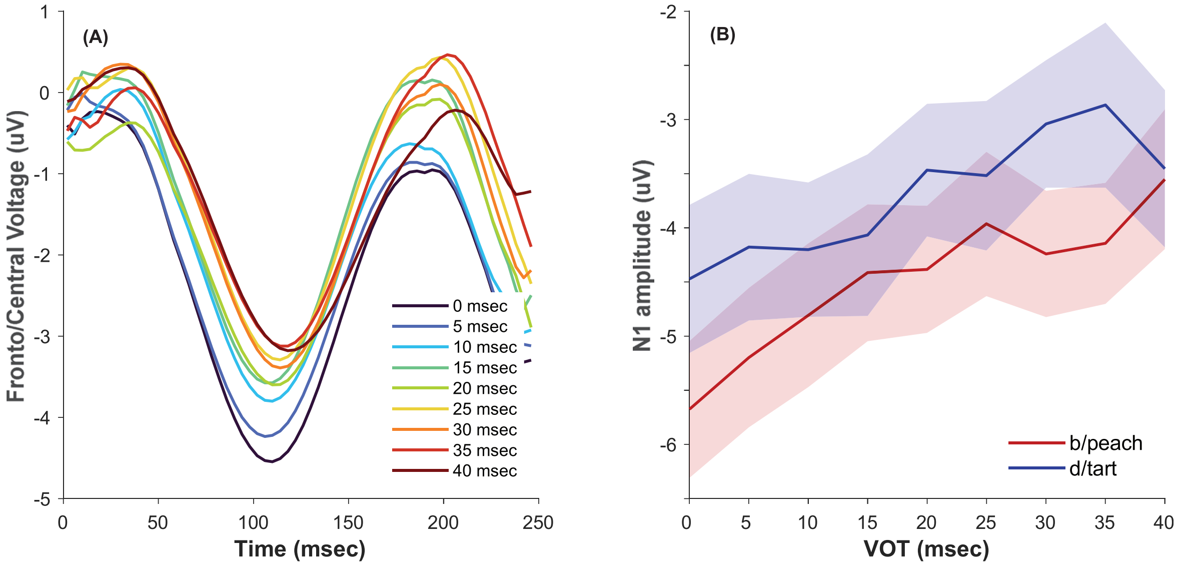


FIG. 7. The results from Toscano *et al.* (2010), showing (A) voltage at front-central channels as a function of time and VOT, where the large negative deflection at 100 msec is the N1 and its depth is related to VOT; and (B) mean N1 amplitude as a function of VOT.

children learn to produce these markers? Similarly, speech pathologists, second language teachers, and dialect coaches must often hear fine-grained differences within a category to help clients achieve more canonical productions. Again, CP would seem to work against this important function of speech perception.

E. Summary

The foregoing discussion challenges CP as an empirical phenomenon at every level. CP does not appear uniformly across speech contrasts, challenging its generality as a model of perception. The discrimination tasks commonly used to establish CP have issues of memory and bias and when these are eliminated, perception is not categorical. Moreover, the model by Pisoni and Tash (1974) (Fig. 5) convincingly argues that if we eliminate the assumption that discrimination *only* assesses the perceptual space, then CP can be observed even if that space is non-categorical (linear). The underlying representation of the continuous space—revealed by continuous ratings and neurophysiological measures—appears linear. Auditory perception must do more than simply yield categories—listening skills, such as talker or dialect recognition, attest to the preservation of within-category detail. There is no reason to hold on to CP anymore and by abandoning it, we may be able to achieve a much more coherent and compelling theoretical account of speech recognition.

IV. SPEECH CATEGORIES ARE NOT DISCRETE (AND THAT IS GOOD)

Categorization is a small part of a system that must ultimately recognize meaningful units—phonemes, words, social identity, and emotion—from a continuous and variable input. CP claims that auditory representations and categories are quasi-discrete. At the time, this aligned with broader ideas about how listeners perceive speech. However, new (and old) thinking suggests that the solution implied by CP is not how listeners approach speech perception.

A. Speech categories are gradient

CP implies that the goal of the perceptual system is to transduce a continuous signal to something resembling a discrete category. Whereas the primary challenges to CP (reviewed above) focus on the perceptual encoding (Fig. 3, perceptual layer), challenges to this broader theoretical account suggest that phoneme categorization and even downstream word recognition are not discrete but gradient, and this is important for efficient processing.

The earliest work on this comes from Miller and Volaitis (1989) and, for a review, Miller (1997). They used a task in which listeners heard tokens from a speech continuum and rated how good of a /p/ each exemplar was. This allowed a visualization of the "structure" of the entire category and revealed that speech categories have a \gradient, prototype-like structure and this entire structure—not just the boundary—was sensitive to contextual factors such as speaking rate [Miller and Volaitis, 1989; Fig. 8(A)]. Moreover, the structure and shape of these goodness ratings roughly align with the distributions of VOTs that listeners experience [Fig. 8(B)]—listeners categorize speech in a way that is sensitive to the statistical distributions of those cues.

It is possible that early representations of speech categories are gradient but transformed into something discrete before accessing higher-level units such as words. This was disconfirmed by work by Andruski *et al.* (1994). They used cross-modal priming to show that small changes in VOT—which did not alter the perceived category—reduced semantic priming, suggesting that gradiency in speech categories is preserved through the lexical level.

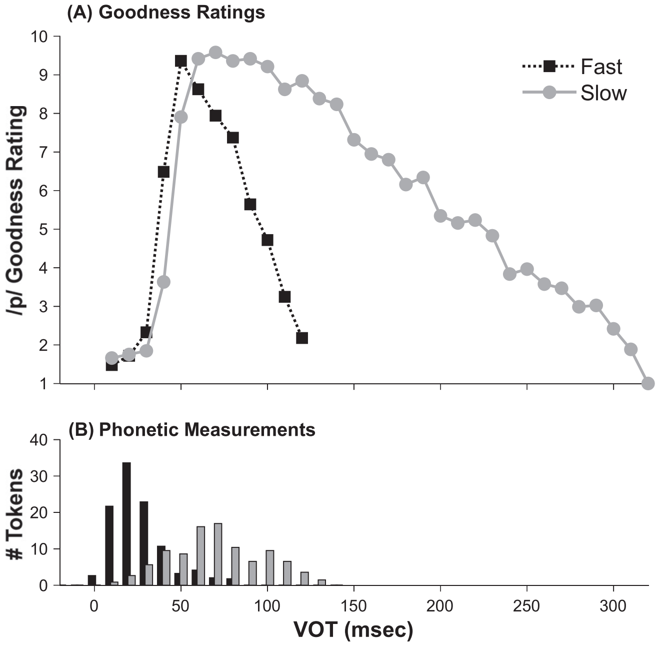


FIG. 8. Typical results of phoneme goodness rating experiments of Miller and colleagues with data estimated from Miller and Volaitis (1989), depicting (A) /p/ goodness as a function of VOT and speaking rate (the length of the following vowel); and (B) phonetic measurements showing the distribution of VOT as a function of speaking rate. Adapted with permission from Miller and Volaitis, Percept. Psychophys. 46(6), 505–512 (1989). Copyright 1989 Springer Nature.

One concern with the priming and goodness rating techniques is that they could represent an average of different types of trials. For example, in the paradigm by Andruski *et al.* (1994), on some trials, *king* may be fully active and fully prime *queen*, but on other trials it is completely inactive and shows no priming. When the VOT is near the boundary, more trials should fall into this latter group. This could attenuate the priming effect, on average, even if the effect on any individual trial was the same across VOTs. To rule out this concern, one needs a technique that queries which category the listener heard on that trial while simultaneously assessing the degree of underlying activation for that category.

A number of studies accomplished this using the visual world paradigm (VWP; McMurray *et al.*, 2008; McMurray *et al.*, 2002). In these studies, listeners heard a token from a speech continuum spanning two words (e.g., *beach* and *peach*) and clicked on the matching picture. This task requires one or more eye movements to plan the response, which can reveal the underlying activation of a word. For example, if the listener heard *beach*, fixations to the *peach* can indicate how active the competing word is. These studies rule out the averaging artifact by analyzing the data relative to each subject's own boundary. Here, −5 msec of VOT then refers to 5 msec from the boundary (toward /b/), and +10 is 10 msec toward /p/. This avoids the possibility that increased looking is driven by variation in the boundary across subjects. Additionally, the analysis of fixations discards trials in which the "incorrect" response (for that side of the continuum) was chosen. This then asks: given that the listener heard a sound that was a fixed distance from their boundary and they ultimately reported a /b/, do competitor fixations (to /p/) vary as a function of VOT?

This approach typically shows a linear effect of VOT on competitor fixations [Figs. 9(A) and 9(B) from McMurray *et al.*, 2002]. As the VOT moves toward the boundary (at relative VOT = 0), there are systematically more fixations to the competitor (see also Kapnoula and McMurray, 2021; McMurray *et al.*, 2008). This approach has also been applied to an ERP index of categorization, the P3 (Kapnoula and McMurray, 2021; Toscano *et al.*, 2010). As with the VWP studies, these analyses accounted for the subject's boundary and their response on each trial and found a stronger P3 for more prototypical VOTs. Such findings are also robust across individuals. Kapnoula and McMurray (2021) correlated individual differences in the gradiency of listeners' speech categorization using the VAS task with VWP and ERP/P3 measures: more gradient VAS responding led to more gradient competitor fixations and P3s. However, all of the listeners were gradient—they just varied in degree.

Goodness ratings, the VWP, and ERPs all suggest that speech categories are gradient—even controlling for the label on each trial and differences in listeners' boundaries. This is indirect evidence against an entirely discrete perceptual representation: if auditory representations were strongly warped, within-category detail would not be available to higher-level lexical processes. This is not strong evidence against weaker forms of CP. It is possible that within-category discrimination is reduced but not lost and sufficient to support a gradient representation. The point is not that these results directly refute CP. Rather, the functional "goal" of speech categorization implied by CP is to ignore irrelevant variation within a category. Instead, these results challenge this framing by showing that listeners systematically track it. Two extensions highlight this fact.

First, this mismatch between the underlying (gradient) activation and the overt (categorical) labeling is highlighted by a recent developmental study. McMurray *et al.* (2018) tested children from 7 to 18 years of age in a variant of the VWP task by McMurray *et al.* (2002). Identification curves (mouse-clicks) showed increasingly steep categorization with development [Fig. 9(C)]. Children appear to become increasingly categorical over-development; in a CP framework, they lose access to fine-grained detail overdevelopment. However, fixations showed the opposite pattern [Fig. 9(D)]: the 7–8-year-olds (red curve) showed essentially no sensitivity to fine-grained detail; this emerged in later age groups. Thus, rather than development leading to the loss of sensitivity, children achieve greater sensitivity to fine-grained detail with age, and this enables sharper categorization.[2] This mismatch between a steepening 2AFC slope but a more gradient underlying representation undercuts a core assumption of CP but also suggests limits to 2AFC categorization tasks (Apfelbaum *et al.*, 2022).

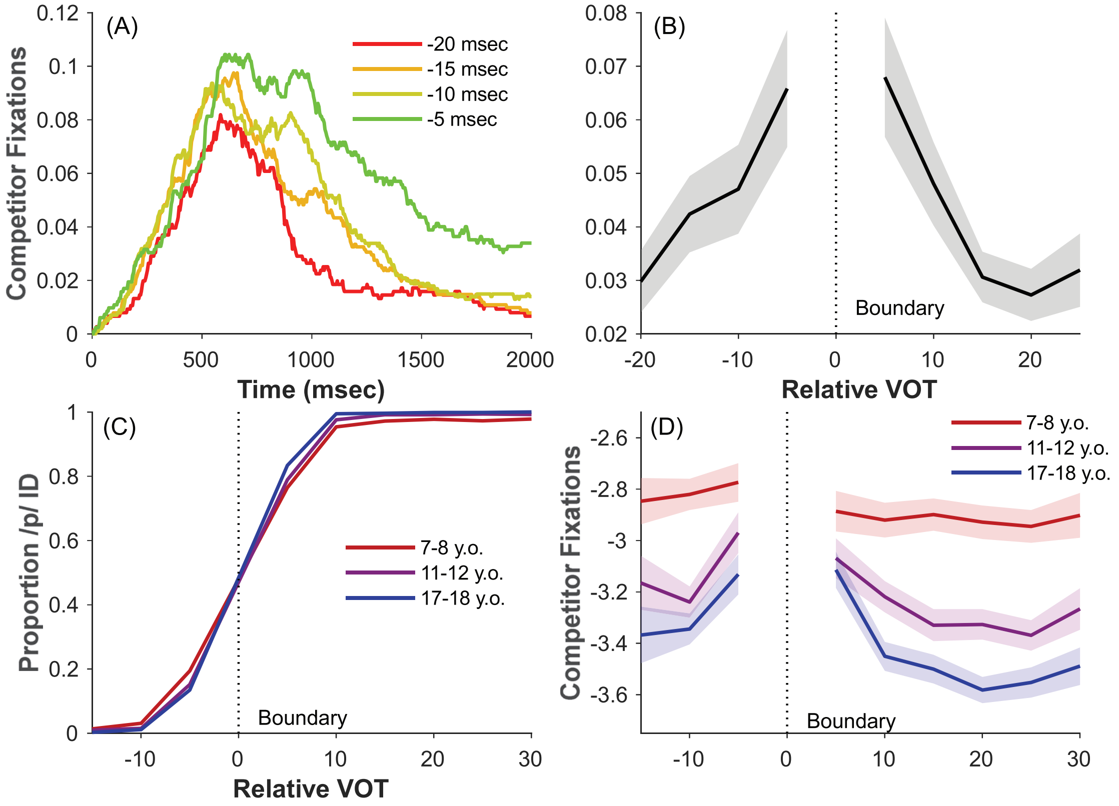


FIG. 9. The results of VWP experiments showing gradient within-category sensitivity to fine-grained detail. (A) Fixations to the competitor are depicted as a function of time and distance from the category for tokens on the voiced side of the continuum from McMurray *et al.* (2002). (B) Average fixation from 200 to 2200 msec as a function of relative VOT in that study are shown. (C) 2AFC labeling function for children is shown in which identification becomes steeper (more categorical) with development. Adapted with permission from McMurray *et al.*, Dev. Psychol. 54(8), 1472–1491 (2018). Copyright 2018 American Psychological Association. (D) In contrast, competitor fixations as a function of relative VOT show increasing sensitivity with development. Data are from McMurray *et al.* (2018).

Second, Myers *et al.* (2009) used functional magnetic resonance imaging (fMRI) to localize the neural basis of speech categorization. They found a highly gradient pattern in STG—an auditory/phonological area, but a more categorical response function in the inferior frontal gyrus, a down-stream language area (see also Toscano *et al.*, 2018). This supports a model closer to that of Fig. 5 in which the system maintains a continuous perceptual representation of the signal (in earlier areas such as STG) and something more abstract and cognitive [in inferior frontal gyrus (IFG)].

B. And gradiency is helpful

Modern theories of speech perception agree that auditory input is represented continuously and activation for categories is gradient. This is true for a wide variety of theories from radically different theoretical bases, including theories from an information integration perspective (McMurray and Jongman, 2011; Nearey, 1997; Oden and Massaro, 1978), connectionism (McClelland and Elman, 1986), Bayesian/ideal observer frameworks (Kleinschmidt and Jaeger, 2015), auditory accounts (Kluender *et al.*, 2003), and exemplar theories (Goldinger, 1998). This is for a good reason—many of these theories posit representations or operations that are simply incompatible with discontinuous auditory input or discrete categories. CP is fundamentally about *reducing access to (within-category) information in the signal* while all of these theories maximize how this information is maintained and used.

It is beyond the scope of this paper to attempt a detailed review of these theories. However, it is important to discuss the general principles, the facts about speech, and empirical results that argue that gradient representations are beneficial for solving the computational problem of speech perception. Several of these phenomena—which were well understand even at times when CP was dominant—are fundamentally inconsistent with CP. That is, if listeners only had a categorical representation, they may not be able to recognize indexical variation, account for context, or learn new dialects (etc). This theoretical inconsistency is a powerful argument against CP and for a gradient alternative.

1. Cue integration

Multiple cues contribute to phonemic percepts (Lisker, 1986; Oden and Massaro, 1978; Repp, 1982). For example, voicing is primarily cued by VOT (in English), but fundamental frequency (F0), F1 frequency, and the length of the subsequent vowel also contribute. These trading relationships often appear as a shift in the boundary along a primary dimension (e.g., VOT) as a function of a secondary cue [Fig. 2(B)]. Although multiple cue integration was always seen as consistent with CP (Repp, 1982; see the supplementary material S3[1]), depending on where and when cue integration occurs relative to categorical warping, CP could hinder the use of multiple cues. That is, to perform these kinds of sensitive boundary shifts, listeners must be able to encode cues with a high degree of precision. If all they heard was /b/ or /p/, it is not clear how this could occur.

Supporting this, several studies have correlated individual differences in multiple-cue use and gradiency using VASs (Kapnoula *et al.*, 2021; Kapnoula *et al.*, 2017; Kong and Edwards, 2011; Ou *et al.*, 2021). Gradiency in VOT has been linked to the degree to which listeners also use F0 for voicing judgements (Kapnoula *et al.*, 2021; Kapnoula *et al.*, 2017; Kong and Edwards, 2011; Ou *et al.*, 2021); and gradiency in tracking formant frequencies has been correlated with the use of duration in vowel categorizations (Ou *et al.*, 2021). Correlations are not observed for some cue combinations, such as fricative spectra and formant transitions, or VOT and vowel length (Kapnoula *et al.*, 2021). However, for at least some cues, gradient representations are linked to better cue integration.

2. Accounting for context

A related phenomenon is context. Boundaries also shift as a function of factors that themselves do not directly cue a phoneme. For example, fricative boundaries shift depending on the gender of the talker (Strand, 1999). Whether or not a phoneme was spoken by a male or female does not directly inform the listener whether a sound is /s/ or /ʃ/. However, sociolinguistic differences lead different talkers to produce these sounds differently, and listeners can do better if they account for this during perception. The problem is that context is not always available with the phoneme in question: for example, the spectral mean of the fricative (indicating /s/ or /ʃ/) arrives before gender information available in the vowel. If the listener made a categorical decision using the bottom-up phonetic cues alone, they would lose some of the sensitivity they need to update this decision later as a function of context.

The C-CuRE (computing cues relative to expectations) model of speech perception illustrates how a non-categorical representation of cues is needed to deal with context (Cole *et al.*, 2010; McMurray and Jongman, 2011). In this model, phonetic cues are encoded continuously but adjusted relative to contextual factors. For example, a spectral mean of 5000 Hz might be recoded as 1000 Hz, which is below what would be expected for a female talker. McMurray and Jongman (2011) showed that a simple classifier model using these adjusted cue values could yield a pattern of performance that closely mirrored that of listeners. Critically, this illustrates an approach to lack of invariance and one in which feedback from phoneme categories (and other things) affects the auditory space *but without the loss of information posited by CP*. For example, Fig. 4(D) offers a visualization of the same fricative cues from Fig. 4(A) but after subtracting out expectations based on the talker and neighboring vowel. To do this, the system must retain as close to a continuous representation as possible to get the most out of this form of processing. Like CP, this model does not posit a veridical representation of the cue: cue values are coded relative to internally generated expectations (e.g., this spectral mean is low for a woman). However, the fundamental operations are different: CP dichotomizes a dimension and arbitrarily eliminates variability, whereas C-CuRE shifts the estimated cue value to more sensitively account for context and increase discriminability (discriminant analysis went from 92% to 96.6% correct). Thus, it is possible to obtain the supposed benefits of CP without throwing away within-category detail.

3. Listeners do more than one thing with any speech cue

A deep challenge to CP is rooted in an observation made by Mermelstein (1978): every speech cue is used for more than one thing (see also Smits, 2001; Whalen, 1992). In English, vowel length serves as a cue to speaking rate; it distinguishes tense and lax vowels, and contributes to syllable-initial and syllable-final voicing. This was a fundamental to the problem of lack of invariance: any cue value is affected by multiple factors, and each factor affects multiple cues.

The problem is that if a listener categorically warps the input for one set of categories, they lose information that could be useful for another. For example, the spectral mean of a fricative is a robust cue for whether it is /s/ or /ʃ/. However, spectral mean also contains information for the upcoming vowel: fricatives before rounded vowels like /u/ have lower spectral means than before unrounded vowels like /i/ (Daniloff and Moll, 1968). Critically, this is a within-category difference: /si/ and /su/ have spectral means within the range of /s/. Consequently, if CP minimized within-category differences, the difference between /si/ and /su/ would be lost and listeners could not use the frication to anticipate an upcoming vowel—as they clearly can (McMurray and Jongman, 2016; Yeni–Komshian and Soli, 1981).

Virtually every cue is affected by multiple factors. CP ignores this to focus on the coding and perception of a limited number of cues for a single phonetic contrast. However, when we consider CP in light of this basic fact about speech perception, it is hard to see how perception can be discontinuous and, yet, listeners could use speech cues for more than one thing.

4. Speech cues do not just encode phonological categories

Relatedly, speech cues reflect more than phoneme categories. VOT and fricative spectra are related to the gender and social identities of the talker (Allen *et al.*, 2003;

Munson, 2007; Zimman, 2017), and vowels carry substantial information about dialect and language community. Cues like F0 are used to signal vocal emotion while also contributing to vowel and consonant identity (Ohde, 1984). People use fine-grained within-category differences to tune their own productions (e.g., to cope with something in their mouth or a missing tooth) or others (e.g., a parent or speech pathologist working with a child). If within-category information is reduced, these judgements become impossible or harder. In fact, active work in speech pathology suggests that more experienced speech pathologists exhibit more gradiency when judging children's productions with the VAS task (Meyer and Munson, 2021). In this view, CP appears to privilege the goal of identifying phonological categories over these other crucial roles of speech. In contrast, a continuous representation of the auditory space coupled to gradient categories would not sacrifice these abilities.

Although these fields have not widely wrestled with the deep implications of these functions for CP (I could find few published papers), these phenomena are widely acknowledged, and my view as an outsider is that fields like sociolinguistics or speech pathology are often just ignoring CP and moving on (though, see Plichta and Preston, 2005, for an elegant demonstraton of non-CP of geographic dialect). Abandoning CP as a core property of speech may permit or help unify models of phoneme recognition with these other functions.

5. Flexibility

A gradient representation may also support perceptual flexibility. Even with multiple cues and compensation, speech input may be ambiguous due to speech errors, an unfamiliar accent, or sheer noise. In these cases, it may be helpful to avoid strong categorical commitments and keep options available in case initial commitments must be revised.

McMurray *et al.* (2009) tested this using a "lexical garden path" paradigm inspired by work in sentence processing. In a word like *parricade*, if the onset sound was ambiguous between /b/ and /p/, the word could be briefly consistent with *barricade* and *parakeet*, and resolution would not occur until late in the word (at -*cade* or -*keet*). Indeed, when listeners heard *parricade* with a VOT of 40 msec, they were initially biased to interpret the input as the beginning of *parakeet* and struggled to revise their decision when -*cade* arrives. In contrast, when the VOT was 30 msec (still a /p/), recovery was faster because *barricade* was more active. If listeners had been categorical, there would not be an ongoing activation of *barricade* for either VOT, and they could not show this gradient recovery. Most recently, Kapnoula *et al.* (2021) found that participants with a more gradient profile of speech categorization (in the VAS task) were better able to recover from these garden paths. This supports the broader argument—more gradiency leads to more flexibility—while raising questions about individual differences that are not accounted for by current models.

One could argue that such effects could be handled by a system that was initially gradient but rapidly resolved to more discrete commitment (in the lexical garden-path paradigm, the ambiguity lasted about 250 msec). However, later studies using sentential context show that these benefits can last a second or more (Brown-Schmidt and Toscano, 2017; Falandays *et al.*, 2020). Moreover, ERP work by Sarrett *et al.* (2020) used the N1/VOT paradigm of Toscano *et al.* (2010; Fig. 5) to show a linear effect of VOT on the EEG as late as 900 msec after the VOT. Critically, target words were ambiguous and at the end of the sentence—no further information was coming. Thus, participants maintain a gradient representation for quite a while even if they will not need it. This may enable more robust and flexible speech perception.

6. Learning and plasticity

Finally, gradiency may be necessary for learning and adaptation. Infants and adults take advantage of statistical learning mechanisms to acquire the categories of their language (Maye *et al.*, 2002), acquire new categories of a second language (Escudero *et al.*, 2011), and tune existing categories to adapt to novel talkers or contexts (Munson, 2011). These mechanisms work by tracking the frequency of occurrence of individual cue values [e.g., how frequent is a VOT of 20 or 25 msec as in Fig. 6(B)] to identify the number of categories along a dimension as well as their prototypical values and the allowable extent.

While there is debate about the sufficiency of these mechanisms to fully account for first-language speech category acquisition (McMurray, 2022; Schatz *et al.*, 2021), listeners do learn from such statistically structured input (Escudero *et al.*, 2011; Maye *et al.*, 2002). Indeed, gradient categorization (Fig. 6) may reflect the fact that these categories are a product of statistical learning. That makes sense if the goal is to adapt the system to the degree of uncertainty: when there is clear evidence that a given token is a /p/, there is no need to hedge one's bets, but when the current input is from a less frequent region of the space, it may be useful to withhold a strong commitment. This kind of statistical learning is impossible in a system in which representations of speech cues are warped (CP). That is, to track the frequency of specific cue values (e.g., how often a 20 msec VOT occurs), listeners must maintain a continuous representation of the cues.

Standard views of infant development assume a more or less categorical end-state of development (Werker and Curtin, 2005); these models might argue that babies are continuous or gradient long enough for such mechanisms to acquire a more categorical system (the learning account described above). However, even after infants have achieved native-language-like discrimination of voicing (at 4 months old), they discriminate speech in a gradient not categorical manner (McMurray and Aslin, 2005; Miller and

Eimas, 1996). Moreover, distributional learning is clearly operative in adults for second language acquisition or fine-tuning categories to new dialects or talkers. A categorical representation would preclude these mechanisms.

The most compelling link between these kinds of gradient representations and plasticity comes from studies that ask how listeners adapt to variability in phonetic cues (Clayards *et al.*, 2008; Theodore and Monto, 2019). In these studies, listeners perform a standard identification task with stimuli from a continuum. For some subjects, the distribution of tokens across trials was tightly clustered around two prototype values (e.g., VOTs of 0 and 40 msec) with little variation; for others, the distribution had a much higher variance, creating trial-to-trial uncertainty. The subjects with wider distributions showed shallower identification slopes and more gradient eye-movement responses. Thus, fine-grained continuous detail is used not only to update or learn categories but to manage uncertainty.

C. Summary

There is not strong evidence that fundamental auditory representations are warped by categories (Gerrits and Schouten, 2004; Schouten *et al.*, 2003) and mounting evidence that they are not (Toscano *et al.*, 2018; Toscano *et al.*, 2010). Speech categories are gradient, and gradiency is preserved as far as lexical processing (Andruski *et al.*, 1994; McMurray *et al.*, 2002) and for considerable time (Falandays *et al.*, 2020; Sarrett *et al.*, 2020). This kind of gradiency is not inconsistent with weaker forms of CP—provided that some within-category sensitivity is available, listeners can exhibit a gradient category. However, that is not the point of this argument.

CP must be embedded in a wider theory of speech processing that enables core perceptual functions such as cue integration, context sensitivity, and the like. I have argued that continuous representations of speech cues and gradient representations of categories are necessary for many such operations, and a categorical encoding may be a hindrance. Gradiency may help integrate multiple cues and contextual factors; it is necessary to deal with the fact that speech cues serve multiple goals—multiple phonemes must be recognized from the same segment of speech, and listeners use it to tune articulations and make sociolinguistic judgements. Gradiency helps maintain flexibility in the face of uncertainty, and it is an essential ingredient—and product of—perceptual learning. The essence of CP is a reduction of within-category phonetic detail; however, these functions require the opposite: preserving and using such detail. While gradiency does not solve the problems of lack of invariance alone (Kapnoula and McMurray, 2021), the benefits of gradient processing are too wide-ranging to ignore.

V. CAN SOME FORM OF CP BE SAVED? AND SHOULD WE?

Can some form of CP be salvaged? Given the state of the evidence, CP cannot be fully ruled out. Yet, why keep it?

Clearly, there is within-category sensitivity and some level of the system that encodes speech cues linearly. However, is the evidence for gradiency inconsistent with CP? In fact, a gradient category representation can live on top of (or in parallel to) a partially warped cue encoding as long as sufficient within-category detail is available to support it. So, none of the work on gradient categories rules out weaker forms of CP.

Yet, is there evidence for some form of auditory warping? Perhaps there is. It is difficult, at this point, for discrimination tasks to offer unambiguous evidence for CP. However, auditory neuroscience may be interpreted as supporting a hybrid. The prior work showing a partially warped N1 in some listeners (Kapnoula and McMurray, 2021) could be an example of some form of CP (in at least some listeners), or it could be explainable in a Pisoni and Tash (1974) model as the simultaneous contribution of a linear cue encoding and a category on the N1. Moreover, MMN (Phillips, 2001) is well-replicated as a tool for eliciting CP. This is harder to dismiss. This component is pre-attentive (it can be elicited during sleep), which rules out explicit strategies and suggests a low-level phenomena. However, it also may be the product of an implicit simultaneous comparison of the baseline and target at the cue and category levels (Fig. 5) as long as categorization is also rapid and pre-attentive. Moreover, the need for many repetitions of a baseline stimulus raises the possibility of learning/adaptation and makes it more difficult to interpret it as an unambiguous measure of cue encoding. Critically, we need more work understanding what processes drive these neural responses to comprehend the degree to which they challenge a gradient model.

Even as we can not fully rule out CP, there is not much unambiguous positive evidence to keep it. The alternative theoretical model is compelling. CP makes a strong claim that something that supposedly occurs prior to categorization (auditory encoding) is warped by categorization. It requires strong evidence. However, the converse, that categorization carves up continuous cues, is not controversial. Given the dearth of hard evidence and to the extent that many results can be explained by mere categorization, maybe we should just posit that.

Given the evidence for some level of continuous encoding, the real question is not just is there or is there not some additional warping somewhere else in the system. The real question is whether listeners (and downstream speech processes) have *access* to a veridical representation of the speech signal (and they clearly do) and which kind of representation is more important for language. Even the original conception of CP would be perfectly happy with a veridical representation that was transformed into a warped representation. However, the deeper implication was that the warped representation was the basis of downstream processes (such as word recognition). In this light, the crucial insight is that the non-categorical or gradient representation is the basis of downstream processes like word recognition. Even if you want to admit some kind of warped auditory encoding, given the role of fine-grained detail in most modern theories of speech, is it wise to make CP central to our field? Should it be one of the few things taught about speech perception in introductory cognition classes? Or should it be relegated to a side issue or caveat to prevailing models?

Finally, perhaps I am splitting hairs. Perception is not just auditory encoding—it should include anything that reaches the level of awareness (e.g., categorization). I agree. The arbitrary line between cognition and perception is old-fashioned and out of date. Yet, if we embrace this more holistic view, CP becomes trivial—it just means that people categorize things. Part of the theoretical richness of CP is that it offers clear definitions about levels of the system and counterintuitive interactions among them. However, if we expand our notion of perception to include categorization, when we argue that people "perceive" items within a category as more similar, does this not just entail the trivial point that they categorize them more similarly or judge them more similarly? What is the added insight of invoking CP?

VI. CP IS DANGEROUS FOR OTHER FIELDS

The foregoing discussion has focused on whether speech is perceived categorically or gradiently, which are issues internal to speech categorization. However, CP has been impactful in several fields outside of the narrow community that has worked on this problem in neurotypical adults with typical hearing listening to speech in their native language. This includes work on the nature of speech categories in *development* and *communicative disorders* and *bilingualism/L2 speech perception*. In these domains, the question is not whether speech is perceived categorically. Rather, studies have assumed CP in order to understand the nature of perception in different populations. With the fall of CP, where does that leave these endeavors? Has it led to incorrect conclusions? Moreover, CP has informed work in *other domains of perception*. Here, CP has supported claims about modularity or linguistic relativity but often without an attempt to wrestle with the nature of the discrimination tasks.

A. Infancy

There is near consensus in the developmental literature on the development of speech categorization (Kuhl, 2004; Werker and Curtin, 2005; but, see McMurray, 2022). Infants "start" with the ability to discriminate many speech contrasts of the world's languages, and over the first 12–18 months of age, discrimination narrows to only the sounds of their language. This appears to accord with CP: auditory representations warp over development to support categorization. How do we synthesize this with more modern, gradient views (cf. McMurray, 2022)?

1. What is really known about infant speech categorization?

First, with the demise of CP, we may know less than we think about these developments in infancy. There are few ways to instantiate an identification or categorization task with infants (though, see Albareda-Castellot *et al.*, 2011), consequently, almost all methods focus on discrimination. Typically, these methods which test discrimination by repeating a single baseline stimulus and changing it to either a within-category variant or between-category variant (with the same physical distance, e.g., Eimas *et al.*, 1971). The infant's response to this novelty is then interpreted as indexing discrimination. By assuming CP, one can make inferences about categories: if discrimination is predicted by identification (CP), a failure to discriminate two tokens indicates categorization, whereas successful discrimination implies two categories. If we cannot assume CP, these data cannot be directly interpreted in terms of categorization. In fact, two studies show evidence of within-category discrimination in infancy (McMurray and Aslin, 2005; Miller and Eimas, 1996) and a systematic review shows evidence for gradient sensitivity when data were pooled across studies that individually reported CP (Galle and McMurray, 2014). Thus, it is not clear that the assumption of CP—necessary for concluding anything about categorization from these tasks—holds.

However, even if we take these studies at face value, there are alternative explanations of these results that do not require CP. As these are discrimination measures, the framing of Pisoni and Tash (1974; Fig. 5) is relevant. In this model, infants have categories (or sets of micro-categories: Schatz *et al.*, 2021) alongside a veridical auditory/perceptual representation and both contribute to the response. Again, when auditory and category representations differ, the discrimination response (e.g., dishabituation) should be larger. Even without warped auditory encoding, the presence of some kind of category on top of the perceptual encoding may be sufficient to drive what appears to be increasing sensitivity to the native language. Thus, under this model, developmental differences are driven by the strength of the category representations which gradually come to dominate discrimination, even if auditory representations are unchanged. This account awaits explicit tests (ERPs may be promising in this regard). However, it reinforces the idea that as long as we posit the uncontroversial idea that auditory and category representations shape infant responses—and certainly both would be available after many repetitions of a baseline stimuli—we should observe CP-like effects without a discontinuous warping.

Supporting this view, meta-analyses (Galle and McMurray, 2014; Tsuji and Cristia, 2014) and individual studies (Kuhl *et al.*, 2006) suggest that the growth of between-category sensitivity may be a dominant pattern of change, rather than the loss of within-category sensitivity. While it is no longer entirely clear that phoneme categories are even acquired during infancy (Feldman *et al.*, 2021;

McMurray, 2022; Schatz *et al.*, 2021), this argument demonstrates how infant results—such as adult discrimination—provide little strong evidence for perceptual warping (CP).

2. Perceptual narrowing

The dominant story of early development is a form of perceptual narrowing (Maurer and Werker, 2014)—infants lose the ability to discriminate things they do not need. This is consistent with the theory inspired by CP that the goal of perception is to filter out unnecessary variation. However, this view is no longer tenable. Instead, it appears that development may be more concerned with more accurately characterizing the statistical structure of speech cues and learning the various factors that give rise to the observed data. For some sounds (e.g., non-native contrasts), this may mean narrowing; but others may need to be enriched, and infants may need to gain sensitivity to contextual factors such as talker. In fact, our work on later periods of development [Fig. 9(D); McMurray *et al.*, 2018] suggests that older children *gain* abilities to encode fine-grained, gradient detail—the exact opposite of narrowing. In short, if speech perception is about harnessing variability to enable flexible behavior, we may need a new metaphor.

These arguments suggest the need to think differently about what infants know, what they are trying to achieve developmentally, and how they get there. This may require moving beyond categories (which are difficult to measure in infancy). For example, Feldman *et al.* (2021) argue that the primary achievement of infancy is organizing the perceptual space (the middle layer of Fig. 3) and true categorization may not come until later childhood, which brings access to more words, articulation, and richer social cues (see also McMurray, 2022).

B. Development and disorders

CP has also been instrumental in understanding development and communication disorders in older children. In typical development, classic identification-from-a-continuum paradigms demonstrate changes in phonetic categorization over the school-age years and later [e.g., Fig. 9(C); Bernstein, 1983; Hazan and Barrett, 2000; McMurray *et al.*, 2018; Nittrouer, 2002; Slawinski and Fitzgerald, 1998]. A large body of work has used similar tasks in people with dyslexia, developmental language disorder, or brain damage (e.g., Dial *et al.*, 2019; Robertson *et al.*, 2009; Serniclaes *et al.*, 2004; Werker and Tees, 1987). These studies typically find that younger or impaired listeners show differences in the identification curve relative to typical or older listeners [Fig. 2(C)] with either a shallower slope (gray curve) or asymptotes that do not reach zero or one. The question is, what does this mean?

The typical CP-inspired interpretation is that listeners strive for discrete categorization. Consequently, a shallower slope derives from noise, specifically, in encoding the cue. For example, if a VOT of 15 msec (a /b/) was misheard as 20 msec (a /p/) on some trials, this would cross the boundary and result in a different identification. However, if a VOT of 0 msec was misheard as 5 msec, it would still be a /b/ and yield no difference. A shallower amplitude may derive from noise at category level (all of the /b/'s are occasionally miscategorized as /p/'s). Thus, assuming CP, these patterns can be clearly interpreted: a shallow slope indexes noisy cue encoding, and a reduced amplitude indexes category-level differences.

This logic does not hold if listeners strive for gradient categories. The mapping between a gradient underlying category and 2AFC performance is ambiguous. A shallower slope could indicate a more gradient category—a rational response to uncertainty (Clayards *et al.*, 2008). However, it could also reflect a noisier system (the model assumed by CP). Both could be present in any sample: shallower slopes in some children indicate greater noise, but in others it indicates a useful adaptation. The 2AFC task cannot distinguish these possibilities as it is unclear how listeners map an underlying gradient representation to a discrete response: if a given token is heard as 60% /b/-like, do they match this probability (choosing /b/ 60% of the time), or do they always choose /b/ (winner take all; Nearey and Hogan, 1986)? As argued by Apfelbaum *et al.* (2022), continuous tasks like the VAS task may bypass some of this to uncertainty to precisely characterize any differences.

A second issue is paradigms that seek to understand cue encoding using discrimination tasks; for example, Serniclaes *et al.* (2004) have argued dyslexia can be linked to an inability to sufficiently ignore within-category detail, and Robertson *et al.* (2009) have argued that children with developmental language disorder show impaired between-category discrimination. However, given the arguments against CP at theoretical and methodological levels, it is not clear what these claims mean. If category and auditory levels contribute to categorization, perhaps these differences simply reflect poorer categorization. Of course, discrimination tasks tap working memory, cognitive control, and phonological processing skills that co-develop with speech perception and may also be impaired in communicative disorders. Perhaps, differences in discrimination do not reflect perception or categorization at all but differences in processes external to speech categorization. Given these issues, it may be better to explore other methods such as continuous VAS tasks, eye-tracking in the VWP, or ERPs.

Beyond methods, the claim of CP is that steep categorization and the suppression of within-category detail is the goal of development. This is what children should learn to do and what impaired listeners do not do as effectively. More modern thinking in speech perception suggests that this is not the case. Gradiency is important—for speech categorization and other functions such as articulatory control and sociolinguistic processing. It is functionally beneficial for listeners and a crucial avenue by which people adapt to communicative impairments. That is, it is something that development attempts to achieve.

C. Bilingualism and L2 acquisition

Work on bilingualism and second language (L2) acquisition has also been motivated by CP. Without attempting a comprehensive review, the demise of CP has two implications.

1. *Critical or sensitive periods*

First, there is a presumption of critical or sensitive periods for L2 acquisition. This is motivated in part by work examining proficiency as a function of when a learner begins L2 acquisition (Johnson and Newport, 1989). However, this work has been challenged by larger-scale studies that show a much bigger sensitive window (through age 18 years old) when we account for not only when L2 acquisition begins but also for the fact that language learning is likely to be protracted (Hartshorne *et al.*, 2018).

The other argument for a sensitive period is the apparent rapid emergence of L1 speech categories in infancy (Werker and Curtin, 2005), coupled with the struggles that adult L2 learners show when acquiring categories (Strange and Shafer, 2008). If infancy is a special time for speech category learning, this can explain these difficulties (Werker and Hensch, 2015). There are two problems with this logic. First, as I described, without the assumptions of CP, it is challenging to conclude that infants truly have acquired speech categories so rapidly. While they are certainly attuning to their language, they may be doing other things (Feldman *et al.*, 2021; McMurray, 2022; Schatz *et al.*, 2021), and development may be slow (Hazan and Barrett, 2000; McMurray *et al.*, 2018), supporting the analysis by Hartshorne *et al.*(2018). Second, infant methods are radically different than those used with adults; a typical infant study may present infants with multiple tokens of the two sounds to be discriminated, and any difference in listening time is used to support discrimination. In contrast, adults get one token/trial. If adult L2 learners were tested with the relatively looser infant procedures, would they look as good? This challenges the basic assumption that L2 learning adults struggle, whereas L1 learning infants "get it." Both of these arguments undercut the premise of a critical period and call for more rigorous and direct empirical evaluation of this construct (Fuhrmeister *et al.*, 2020) as well as new ways of thinking about critical periods (Thiessen *et al.*, 2016).

2. *What is the goal of bilingual and L2 speech perception?*

Most work on bilingual or L2 speech perception is implicitly framed around CP, which restricts our understanding in two ways. First, as with children and impaired populations, it is not warranted to presume that a sharp categorical boundary is desirable. In a multilingual environment, *gradiency may be even more helpful* (than for monolinguals). When confronted with multiple languages (potentially changing from moment to moment), gradiency may be needed for listeners to be more flexible and incorporate context (about the current language). Thus, the question should not be how listeners attain sharp boundaries but how they attain flexibility.

Second, CP emphasizes boundaries. In a bilingual situation that frames questions such as whether a Spanish/English bilingual puts their VOT boundary at 20msec (English) or 0 msec (Spanish) and whether they can shift it sensitively. Yet, no serious model of speech perception assumes boundaries. Instead, prototypes (e.g., Fig. 8; or a functional variant) are the norm. Crucially, *prototypes need not be mutually exclusive*. A listener could have an English /b/ category centered at 5 msec and a Spanish /b/ category centered at −60 msec. Thus, moving beyond CP may pose new questions for work on bilingualism and L2 acquisition.

D. Other domains of cognition and perception

The extension of CP to domains beyond speech has been active, and CP has played an important role in debates about linguistic relativity (Franklin *et al.*, 2008), modularity (McKone *et al.*, 2001), and other fundamental issues. While it is beyond the scope of this review to fully address these issues, it is worth a few brief comments. For the most part, this literature makes little contact with the literature on CP within speech. There are exceptions, but often the deeper message is lost. The influential review by Goldstone and Hendrickson (2010) discusses some of the speech literature, but it does not question the premise. Even the original demonstration of CP in color (Bornstein and Korda, 1984) explicitly adopts the Pisoni and Tash (1974) model (Fig. 3), but fails to realize its deeper implication that CP may be unfalsifiable.

In fact, the few studies that have addressed this issue outside of speech suggest a story more consistent with non-CP. Hanley and Roberson (2011) retrospectively analyzed several studies of face and color CP and conclude that discrimination in these domains is well-described by the model in which perception is not warped and influences discrimination in parallel with categories (e.g., Pisoni and Tash, 1974). Similarly, Roberson *et al.* (2009) identified a less biased discrimination task for color CP and showed no evidence for CP. While these sorts of issues are still under active debate (e.g., Best and Goldstone, 2019), this suggests some movement in the field.

However, the broader argument here is that CP does not live in isolation from broader theories of speech perception—CP makes claims that the rest of the system must live with. In speech, throwing away fine-grained continuous detail limits the degree to which the system has access to this detail for other things such as integrating cues over time, using the same cue for multiple purposes, or identifying sociolinguistic factors. This sort of theoretical inconsistency is relevant to other domains of cognition. For example, in face perception, "norm-based" coding views (Rhodes and Leopold, 2011) suggest that face perception is seen as dimensional, and frequent experience with classes of faces can shift the mean of this dimension—analogous to C-CuRE's (McMurray and Jongman, 2011) approach to talker compensation in speech.

Critically, this kind of dimensional shifting cannot occur if dimensions are collapsed into discontinuous regions. Relatedly, all of the domains of perception confront the problem identified by Mermelstein (1978): that any dimension is needed for more than one purpose. It may make sense to minimize within-category differences in color if the only goal is accurate labeling. Yet, people must also use subtle aspects of color to judge attractiveness or compensate for ambient lighting. In fact, labeling may be one of the least important goals of color perception! Similarly, if observers ignore variation across exemplars of the same face, how can they use subtle differences in that face to judge emotion or health? One can argue that perhaps CP-like effects emerge in the context of specific tasks, but this is a far cry from a radical warping of perception that is the groundwork of higher order cognition.

E. Summary

In development, disorders, multilingualism, and cognition more broadly, a strong presumption of CP has shaped many things, from the tasks used to assess perception and categorization, to the interpretation of the functional goals of the system. With the demise of CP within speech perception, this has not been questioned. It needs to be questioned. In an influential critique of much of psychology, Meehl (1990) coined the term *derivation chain*, the inferential logic that allows one to generate predictions from a theory to an experiment or measure. In these domains, CP was an essential piece of the derivation chain. However, CP is now a part of the past; it makes problematic theoretical assumptions, and it relies on problematic methods. This undercuts the derivation chains used in these lines of work and calls for new approaches.

VII. WHERE DO WE GO FROM HERE?

CP has been around since the cognitive revolution and before I was born. Despite ample evidence to the contrary, it remains. Despite its inability to address fundamental facts about speech (that speech cues do multiple things), it remains. Despite its inconsistency with modern theories of speech perception, it remains. Why? Where do we go from here? I start by considering why it was (and remains) compelling.

A. What made (and makes) CP so compelling?

When CP was discovered and promulgated, speech science was in a different place. While early work identified many key sources of variability (Delattre *et al.*, 1955), we had not identified the hundreds of phonetic cues we have now, we did not have access to large corpora, and statistical tools for pooling "big data" were not widely used. Speech research was grounded in a small number of cues in simple syllables. The problem was narrowly defined in terms of categorization of a single phonemic contrast from a short segment (much as I have simplified here). In this light, it is easy to see how CP was an attractive solution (see the supplementary material S4[1]).

Let us look at the same problem through a modern lens. What if we had encountered the problem of lack of invariance for the first time now, when we have identified far more phonetic cues, we have access to large corpora, there are tools for integrating dozens of measures, our paradigm examines more variable speakers and listeners, and we have a richer understanding of related processes such as word recognition and indexical perception? Would we have come up with CP? It is not clear that we would have. Historical precedence is not an argument retaining CP.

However, the persistence of CP is more than precedence. CP fits with our meta-expectations about language. Any ordinary person can tell you, there is nothing in between a *bunt* and a *punt*—these are distinct words. For speakers of alphabetic languages, there is nothing between a /b/ and a /p/. Arguably, our over-reliance on the IPA for conceptualizing sound contributes. Fieldwork in linguistics, speech pathology, and developmental work has always emphasized that symbolic representation of a fundamentally continuous signal. CP matches this paradigm. We want to be able to talk about sound in terms of categories, and we even think we hear it in terms of categories. However, this is an illusion—when we look closely with unbiased discrimination tasks or EEG, we see a different story. People like to categorize—whether in terms of statistical significance, gender categories, or color—and are uncomfortable with gradience and nuance. CP seems to take perception, which is quintessentially messy, nuanced, and continuous, and offers a comforting answer that it is simple: categories are just fine.

B. What gives? Why is CP still with us?

The scientific community is generally more comfortable with nuance. The community of researchers working on speech categorization has known for a long time that speech is not perceived categorically. So, how has CP persisted? This is illustrated anecdotally by an experience I had as an early graduate student. When I submitted the first few papers arguing for gradiency (McMurray *et al.*, 2008; McMurray *et al.*, 2002), reviewers argued that my work was setting up a straw man, CP was over, and we should move on. Clearly, the reviewers knew what was up. Yet, this message did not appear to be true. Almost everyone outside of speech categorization was still assuming CP.

That is still true 20 years later. Work is building new theories of CP even in speech (Kronrod *et al.*, 2016) and in broader domains of cognition (Feldman, 2021; Zhang *et al.*, 2021). CP continues to be applied in subdomains of speech such as L2 learning and language/hearing disorders.[3] CP is not going away, even as the community of hardcore speech nerds who work directly on speech categorization knows better. Bearing witness to this is the continual appearance of major reviews and large-scale studies of CP (Goldstone and Hendrickson, 2010; Kronrod *et al.*, 2016) that treat CP as a fundamental fact to be explained. Although these often make contact with the literature reviewed here, they do not appear to question the premise of CP itself.

I do not know how those of us who have worked on CP could make the truth more apparent. There are papers with titles like "*The end of categorical perception as we know it*" (Schouten *et al.*, 2003), "*Categorical results do not imply categorical perception*" (Hary and Massaro, 1982), and my favorite, "*Categorical perception of speech: A largely dead horse, surpassingly well kicked*" (Crowder, 1989). Could speech perception researchers be any more direct? One does not even need to read the papers to know that something is up.[4]

Part of the problem is that CP is no longer just an empirical finding or theory. It is part of the meta-narrative of speech, a scientific meme. It is easy to dismiss individual papers here and there while at the same time, treating them as exceptions to a general rule. I hope what this review has done is to show that there are simply too many inconsistencies to hold on to CP anymore. The question is where to go from here (see Box 1 for thoughts).

C. Science communication

Part of the solution must be better communication. There are few, if any, reviews of CP in the speech literature (broadly construed) that wrestle with the more modern conception. Speech researchers of all sub-domains must take it upon themselves to communicate outside of their fields. We can not expect textbook authors and people outside of our field to figure it out from more highly technical papers in the *Journal of the Acoustical Society of America* (yes, I get the irony). If my review was overly pointed, it was for this purpose—subtlety will not help make a dent in the rapid growth of CP (Fig. 1). In fact, even the basic definitions are not making it out of our field. Many people use "CP" to refer to any task in which people identify tokens from a continuum, and they refer to a sigmoidal identification function as a CP curve. This is done regardless of whether discrimination is measured and the researcher is making claims about perception. How do so many papers that mischaracterize or misapply CP continue to be published? Have speech categorization researchers agreed to just let this pass? Are we not reviewing these papers?

A critical issue is teaching. I recently surveyed most of the major psycholinguistics textbooks and many phonetics texts. All of these treat CP uncritically. I suspect it is the same in other fields such as speech and hearing science or phonetics (if speech categorization is covered at all). Why? The easy response is that the alternative is too complicated for undergraduates. However, this sells our students short. The debates around CP are an excellent opportunity to teach how science can be self-correcting, and theories of speech perception must account for fundamental facts about speech. My suspicion is that some professors are making end runs around the texts, to teach CP correctly with innovative demonstrations and tools. If so, these should be shared. Too many texts treat speech perception as an interesting set of phenomena for students to share with their parents: CP, the McGurk effect, duplex perception, and the cocktail party effect. This is important for raising interest, but it does not get our students closer to a deeper understanding of how speech and language works. The gradient alternative is not too complex: use multiple cues, engage in simple forms of compensation, maintain a prototype structure to be flexible, and constantly learn. This is a message we can tell clearly.

D. What do we do with previous results related to CP?

A question now is what to make of the voluminous findings that report the characteristic profiles of between- and within-category discrimination that constitute CP. It is tempting to dismiss them out of hand, but that is wrong. The data are real, even if the premise is problematic. We need a framework for understanding these results and making sense of differences in CP-like behavior across conditions and groups.

Clearly, the wrong conclusion is that perception is warped by the presence of categories. This is not tenable anymore. However, a not unreasonable claim—following Fig. 5—may be that once CP (empirically) is observed, categorization must be pretty robust (fast and strong) to contribute to discrimination. Weakly represented categories do not impact discrimination. For example, when CP is not observed in a L2 (Miyawaki *et al.*, 1975), listeners may simply have categories that are represented or accessed less robustly. If we restrict ourselves to the idea that CP reflects robust categorization, this seems reasonable. However, perhaps we should just call it categorization (or perception of categories) and not CP. If it is just categorization, should we skip discrimination tasks altogether? There are more informative ways to assess categorization.

The alternative is to restrict the domain of interpretation. CP is not a paradigm for understanding perception, but it is useful for thinking about how people perform discrimination judgements. Discrimination is an interesting human behavior and arguably important in its own right. Maybe CP is a way of probing the influence of higher-level knowledge on these judgements. If that is the question, perhaps *AX* and *ABX* tasks serve an important role. Again, this is defensible, but again it is not quite CP. Perhaps this is biased discrimination?

In both approaches, the key is to have a clear understanding of what we are actually studying in CP paradigms so we know what can be concluded. What does not seem reasonable to me is to stretch the concept of CP in every conceivable direction to try to preserve the name.

E. Methods?

The foregoing discussion clearly points to the need for a stronger understanding of methods. The limits of discrimination tasks are clearly highlighted (Pisoni and Lazarus, 1974; Schouten *et al.*, 2003), and the properties and concerns about discrimination tasks should be well known in any field that relies on them. Yet, the 2AFC task is not without its issues either.

As the brief discussion about development reveals, there are limits of interpretation to this task: is a shallow slope a natural reflection of a gradient system or the result of a noisy but categorical system? In fact, even if listeners have an underlyingly gradient boundary, they could still show a steep slope if their response on each trial always reflects the more likely of the two alternatives (Nearey and Hogan, 1986). We simply do not understand the derivation chain (Meehl, 1990) linking an underlying category structure to behavior in this simple task. Even if we did, this task may simply be underdetermined. More sophisticated measures are needed. However, these need not be technically sophisticated (e.g., eye-tracking or ERPs). A history of work (Kapnoula *et al.*, 2017; Massaro and Cohen, 1983; Miller, 1997) suggests that even something as simple as a continuous rating task (Fig. 6) can dramatically open new possibilities for understanding categorization (Apfelbaum *et al.*, 2022). More sophisticated tools, such as eye-tracking and ERPs, are not going to be a panacea. The most common ERP measure of speech, for example, the MMN (see Phillips, 2001, for an example applied to CP), is, in fact, a measure of discrimination with all of the caveats to interpretation that apply to any discrimination task. Techniques like ERPs and the VWP are useful, but they need to be deployed in ways that are linked to the underlying cognitive and perceptual operations with understanding of their own derivation chains. For example, the ability to condition analyses of ERPs or eye movements on the overt response offers the unique ability to ensure that one is examining within-category variation, or the fact that the N1 is tied to early auditory cortical regions offers an ability to isolate processing of continuous acoustic encoding.

CP also introduced to us the notion of a speech continuum. This is arguably the most valuable and defensible methodological innovation. Nothing in the foregoing review invalidates this approach, and the fact that it can span an ambiguous region can highlight subtle but theoretically meaningful effects that cannot be seen elsewhere. However, we can improve it. It would be helpful to standardize how continua are reported and used (see the supplementary material S1[1]). This can help ground arbitrary step numbers to real acoustic measurements. Such standardization can help compare effects across different contrasts and could be crucial for meta-analytic approaches. However, as tools for constructing continua become better and easier, we must be careful that continua are phonetically realistic—that they span neighboring phonemes (not passing through a third phoneme or dead space) and manipulate cues in ways that reflect variation that real articulators are capable of.

Yet, this is not the only way to study speech (note to self). In the era of big data and large-scale statistical models, one can (for example) simply record dozens of naturalistic exemplars, measure their phonetic properties, and give them (unmanipulated) to listeners to categorize. Then, logistic models can sort out which phonetic properties are contributing to a given percept or how that changes across experimental conditions (McMurray and Jongman, 2011). This could provide a powerful—and phonetically grounded—complement to continua.

The last methodological concern is the biggest. Many experiments are predicated on some notion of what kind of processing is typical, efficient, optimal, or ideal. This is perhaps the most serious impact of CP by biasing researchers to expect sharp categorization and poor within-category discrimination as the ideal. Instead, modern theories of speech suggest a system that is much more gradient, flexible, and complex, and our thinking needs to adapt.

F. Conclusions

It is possible to work around many of the points raised here individually. I have argued that CP is incompatible with the kind of rich compensation needed to account for talker and coarticulatory variation. An easy counterargument that could preserve CP would be that listeners do all of this *before* they do CP (or maybe these processes are how one becomes categorical). However, if that is the case, what is the point of adding CP to the system (and in that case, is this really perception)? Attempting to salvage CP in this way ignores the fundamental claim of CP that the system is attempting to suppress variation. One could also attempt to account for the myriad discrimination results by claiming that there is an early categorical representation alongside a linear representation. Yet, is that really any different than the Pisoni and Tash (1974) model?

If one contorts oneself, we can salvage CP from many of these attacks. However, at some point, we have to look at the bigger picture: CP as an empirical finding or a theory is just not logically consistent as a whole. It is not a coherent account of the extant data, and its theoretical claims are incommensurate with theories of speech that use fine-grained detail to do useful work for listeners. Given the wealth of empirical evidence against CP and strong theories of speech that either do not require it or conflict with it, CP no longer adds to the greater coherence of our understanding of speech perception. Can we be finished?

ACKNOWLEDGMENTS

The author would like to thank Keith Apfelbaum, Joe Toscano, Joan Sereno, and Allard Jongman for helpful comments on this manuscript; Ethan Kutlu for engaging ideas about multilingualism, and Effie Kapnoula for ideas regarding individual differences. The author would particularly like to thank Ben Munson, whose thoughtful review of this manuscript led to a rethinking of crucial issues around the relationships among fields and the importance of social and other functions of speech. This project was supported by the National Institutes of Health (NIH) Grant Nos. DC008089 and DC000242.

Speech perception in a post-CP world

CP has shaped the field for so long that it is unclear what the field will look like without it. To start that conversation, here are some concrete (but not exhaustive) recommendations.

Science communication

- Textbooks should be revised to reflect the modern view.
- Pithy review articles are needed to communicate to speech-adjacent fields.
- Reviewers need to be more pointed in challenging appropriate uses of CP.

Methods

- Small differences in discrimination tasks can lead to different memory demands and different levels of bias. We cannot conclude CP from discrimination tasks, and they should be used very carefully.
- Forced-choice identification tasks may be problematic when studying the slope of the identification function (e.g., overdevelopment; Apfelbaum *et al.*, 2022). For this issue, a continuous VAS task may be better (Kapnoula *et al.*, 2017; Munson *et al.*, 2017).
- Forced-choice identification tasks are not problematic for assessing boundary shifts (e.g., trading relations and perceptual learning paradigms).
- Measures that allow researchers to simultaneously measure the category label and underlying activation on the same trial (e.g., specific VWP or ERP paradigms; McMurray *et al.*, 2002) are valuable.
- Speech continua are valuable when they are constructed on the basis of phonetic insight (not arbitrary morphing). It would help to standardize them across dimensions.
- Speech continua should be complemented with more naturalistic methods in which actual recordings are measured along multiple dimensions to relate categorization to actual acoustic variance.

Theory and interpretation

- The focus should be on theories of perception and not theories of categorization along a speech continuum;
- We need stronger theories of the problems highlighted by modern ideas: How do listeners achieve flexibility? Where do individual differences come from? How do listeners use fine-grained acoustic variability?
- Theoretical accounts of development, bilingualism, and clinical populations should abandon the premise that the goal of speech perception is discrete categorization and ignoring within-category detail and, instead, ask how these listeners achieve a gradient, flexible, and efficient categorization.

[1] See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0016614 for additional discussion of standardizing speech continua (S1), discrimination tasks (S2), cue integration (S3), and the history of the field (S4).

[2] This mirrors work by Van Hessen and Schouten (1999), which found that increasing the quality of the continuum—enhancing access to fine-grained detail—led to sharper categorization and more robust CP.

[3] As evidence, a Google/Scholar search of "categorical perception" *and* "L2 learning" led to 1170 hits since 2017, and ("categorical perception" *and* "communication disorders") led to 3020 hits!

[4] Also, I have put the titles right in the text to save you the trouble of checking the references.

Albareda-Castellot, B., Pons, F., and Sebastián-Gallés, N. (**2011**). "The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm," Dev. Sci. **14**(2), 395–401.

Allen, J. S., Miller, J. L., and DeSteno, D. (**2003**). "Individual talker differences in voice-onset-time," J. Acoust. Soc. Am. **113**(1), 544–552.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., and Jones, R. S. (**1977**). "Distinctive features, categorical perception, and probability learning: Some applications of a neural model," Psychol. Rev. **84**(5), 413–451.

Andruski, J. E., Blumstein, S. E., and Burton, M. W. (**1994**). "The effect of subphonetic differences on lexical access," Cognition **52**, 163–187.

Apfelbaum, K. S., Kutlu, E., McMurray, B., and Kapnoula, E. (**2022**). "Don't force it! Gradient speech categorization calls for continuous categorization tasks," J. Acoust. Soc. Am. (in press).

Beale, J. M., and Keil, F. C. (**1995**). "Categorical effects in the perception of faces," Cognition **57**(3), 217–239.

Bernstein, L. E. (**1983**). "Perceptual development for labeling words varying in voice onset time and fundamental-frequency," J. Phon. **11**(4), 383–393.

Best, R. M., and Goldstone, R. L. (**2019**). "Bias to (and away from) the extreme: Comparing two models of categorical perception effects," J. Exp. Psychol.: Learn. Mem. Cogn. **45**(7), 1166–1176.

Blumstein, S. E., and Stevens, K. N. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**(4), 1001–1017.

Bornstein, M. H., and Korda, N. O. (**1984**). "Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing," Psychol. Res. **46**(3), 207–222.

Brown-Schmidt, S., and Toscano, J. C. (**2017**). "Gradient acoustic information induces long-lasting referential uncertainty in short discourses," Lang., Cognit. Neurosci. **32**(10), 1211–1228.

Carney, A. E., Widin, G. P., and Viemeister, N. F. (**1977**). "Non categorical perception of stop consonants differing in VOT," J. Acoust. Soc. Am. **62**, 961–970.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (**2010**). "Categorical speech representation in the superior temporal gyrus," Nat. Neurosci. **13**(11), 1428–1432.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (**2008**). "Perception of speech reflects optimal use of probabilistic speech cues," Cognition **108**(3), 804–809.

Cole, J. S., Linebaugh, G., Munson, C., and McMurray, B. (**2010**). "Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach," J. Phon. **38**(2), 167–184.

Crowder, R. G. (**1989**). "Categorical perception of speech: A largely dead horse, surpassingly well kicked," Behav. Brain Sci. **12**(04), 760–760.

Cutting, J. E., and Rosner, B. S. (**1974**). "Categories and boundaries in speech and music," Percept. Psychophys. **16**(3), 564–570.

Daniloff, R., and Moll, K. (**1968**). "Coarticulation of lip rounding," J. Speech. Lang. Hear. Res. **11**(4), 707–721.

Delattre, P., Liberman, A. M., and Cooper, F. S. (**1955**). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. **27**, 769–773.

Dial, H. R., McMurray, B., and Martin, R. C. (**2019**). "Lexical processing depends on sublexical processing: Evidence from the visual world paradigm and aphasia," Atten. Percept. Psychophys. **81**(4), 1047–1064.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). "Speech perception in infants," Science **171**(3968), 303–306.

Escudero, P., Benders, T., and Wanrooij, K. (**2011**). "Enhanced bimodal distributions facilitate the learning of second language vowels," J. Acoust. Soc. Am. **130**(4), EL206–EL212.

Falandays, J. B., Brown-Schmidt, S., and Toscano, J. C. (**2020**). "Long-lasting gradient activation of referents during spoken language processing," J. Mem. Lang. **112**, 104088.

Feldman, J. (**2021**). "Mutual information and categorical perception," Psychol. Sci. **32**(8), 1298–1310.

Feldman, N. H., Goldwater, S., Dupoux, E., and Schatz, T. (**2021**). "Do infants really learn phonetic categories?," Open Mind **5**, 113–131.

Francis, A. L., Ciocca, V., and Chit Ng, B. K. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**(7), 1029–1044.

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., and Davies, I. R. L. (**2008**). "Lateralization of categorical perception of color changes with color term acquisition," Proc. Natl. Acad. Sci. U.S.A. **105**(47), 18221–18225.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (**2001**). "Categorical representation of visual stimuli in the primate prefrontal cortex," Science **291**(5502), 312–316.

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (**1962**). "The identification and discrimination of synthetic vowels," Lang. Speech **5**, 171–189.

Fuhrmeister, P., Schlemmer, B., and Myers, E. B. (**2020**). "Adults show initial advantages over children in learning difficult nonnative speech sounds," J. Speech. Lang. Hear. Res. **63**(8), 2667–2679.

Gaißert, N., Waterkamp, S., Fleming, R. W., and Bülthoff, I. (**2012**). "Haptic categorical perception of shape," PLoS One **7**(8), e43062.

Galle, M. E., and McMurray, B. (**2014**). "The development of voicing categories: A meta-analysis of 40 years of infant research," Psychon. Bull. Rev. **21**(4), 884–906.

Gerrits, E., and Schouten, M. E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**(3), 363–376.

Getz, L. M., and Toscano, J. C. (**2021**). "The time-course of speech perception revealed by temporally-sensitive neural measures," Wiley Interdiscip. Rev. Cogn. Sci. **12**(2), e1541.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goldstone, R. L., and Hendrickson, A. T. (**2010**). "Categorical perception," WIRES Cognit. Sci. **1**(1), 69–78.

Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (**2001**). "Altering object representations through category learning," Cognition **78**, 27–43.

Green, P. A., Brandley, N. C., and Nowicki, S. (**2020**). "Categorical perception in animal communication and decision-making," Behav. Ecol. **31**(4), 859–867.

Guenther, F. H., and Gjaja, M. (**1996**). "The perceptual magnet effect as an emergent property of neural map formation," J. Acoust. Soc. Am. **100**, 1111–1112.

Guenther, F. H., Nieto-Castanon, A., Ghosh, S., and Tourville, J. (**2004**). "Representation of sound categories in auditory cortical maps," J. Speech. Lang. Hear. Res. **47**, 46–57.

Hanley, J. R., and Roberson, D. (**2011**). "Categorical perception effects reflect differences in typicality on within-category trials," Psychon. Bull. Rev. **18**(2), 355–363.

Harnad, S. (**1987**). *Categorical Perception: The Groundwork of Cognition* (Cambridge University Press, Cambridge, UK).

Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (**2018**). "A critical period for second language acquisition: Evidence from 2/3 million English speakers," Cognition **177**, 263–277.

Hary, J. M., and Massaro, D. W. (**1982**). "Categorical results do not imply categorical perception," Percept. Psychophys. **32**(5), 409–418.

Hazan, V., and Barrett, S. (**2000**). "The development of phonemic categorization in children aged 6–12," J. Phon. **28**(4), 377–396.

Healy, A. F., and Repp, B. H. (**1982**). "Context independence and phonetic mediation in categorical perception," J. Exp. Psychol. Hum. Percept. Perform. **8**(1), 68–80.

Hess, U., Adams, R. B., and Kleck, R. E. (**2009**). "The categorical perception of emotions and traits," Social Cognit. **27**(2), 320–326.

Howard, D., Rosen, S., and Broad, V. (**1992**). "Major/minor triad identification and discrimination by musically trained and untrained listeners," Music Percept. **10**(2), 205–220.

Johnson, J. S., and Newport, E. L. (**1989**). "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language," Cognit. Psychol. **21**(1), 60–99.

Kapnoula, E. C., Edwards, J., and McMurray, B. (**2021**). "Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise," J. Exp. Psychol. Hum. Percept. Perform. **47**(4), 578–595.

Kapnoula, E. C., and McMurray, B. (**2021**). "Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking," Brain Lang. **223**, 105031.

Kapnoula, E. C., Winn, M. B., Kong, E., Edwards, J., and McMurray, B. (**2017**). "Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach," J. Exp. Psychol. Hum. Percept. Perform. **43**(9), 1594–1611.

Kleinschmidt, D., and Jaeger, F. (**2015**). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychol. Rev. **122**(2), 148.

Kluender, K. R., Coady, J., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**(1), 59–69.

Knight, F. L. C., Longo, M. R., and Bremner, A. J. (**2014**). "Categorical perception of tactile distance," Cognition **131**(2), 254–262.

Kong, E. J., and Edwards, J. (**2011**). "Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking," in *Proceedings of the XVIIth International Congress of Phonetic Sciences*, Hong Kong.

Kronrod, Y., Coppess, E., and Feldman, N. H. (**2016**). "A unified account of categorical effects in phonetic perception," Psychon. Bull. Rev. **23**(6), 1681–1712.

Kuhl, P. K. (**2004**). "Early language acquisition: Cracking the speech code," Nat. Rev. Neurosci. **5**(11), 831–843.

Kuhl, P. K., and Miller, J. D. (**1975**). "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," Science **190**(4209), 69–72.

Kuhl, P. K., Stevens, E. H., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (**2006**). "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months," Dev. Sci. **9**, F13–F21.

Lachlan, R. F., and Nowicki, S. (**2015**). "Context-dependent categorical perception in a songbird," Proc. Natl. Acad. Sci. U.S.A. **112**(6), 1892–1897.

Levin, D. T., and Beale, J. M. (**2000**). "Categorical perception occurs in newly learned faces, other-race faces, and inverted faces," Percept. Psychophys. **62**(2), 386–401.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**(6), 431–461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**(5), 358–368.

Liberman, A. M., Harris, K. S., Kinney, J., and Lane, H. (**1961**). "The discrimination of relative onset-time of the components of certain speech and non-speech patterns," J. Exp. Psychol. **61**, 379–388.

Lisker, L. (**1986**). " 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang. Speech **29**(1), 3–11.

Lupyan, G. (**2012**). "Linguistically modulated perception and cognition: The label-feedback hypothesis," Front. Psychol. **3**, 54.

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). "The psychophysics of categorical perception," Psychol. Rev. **84**(5), 452–471.

Massaro, D. W., and Cohen, M. M. (**1983**). "Categorical or continuous speech perception: A new test," Speech Commun. **2**, 15–35.

Massaro, D. W., and Hary, J. M. (**1984**). "Categorical results, categorical perception, and hindsight," Percept. Psychophys. **35**(6), 586–588.

Maurer, D., and Werker, J. F. (**2014**). "Perceptual narrowing during infancy: A comparison of language and faces," Dev. Psychobiol. **56**(2), 154–178.

May, B., Moody, D. B., and Stebbins, W. C. (**1989**). "Categorical perception of conspecific communication sounds by Japanese macaques, *Macaca fuscata*," J. Acoust. Soc. Am. **85**(2), 837–847.

Maye, J., Werker, J. F., and Gerken, L. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**, B101–B111.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**(1), 1–86.

McKone, E., Martini, P., and Nakayama, K. (**2001**). "Categorical perception of face identity in noise isolates configural processing," J. Exp. Psychol. Hum. Percept. Perform. **27**(3), 573–599.

McMurray, B. (**2022**). "The acquisition of speech categories: Beyond perceptual narrowing, beyond unsupervised learning and beyond infancy," Lang., Cognit. Neurosci. (published online).

McMurray, B., and Aslin, R. N. (**2005**). "Infants are sensitive to within-category variation in speech perception," Cognition **95**(2), B15–B26.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., and Subik, D. (**2008**). "Gradient sensitivity to within-category variation in words and syllables," J. Exp. Psychol. Hum. Percept. Perform. **34**(6), 1609–1631.

McMurray, B., Danelz, A., Rigler, H., and Seedorff, M. (**2018**). "Speech categorization develops slowly through adolescence," Dev. Psychol. **54**(8), 1472–1491.

McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**(2), 219–246.

McMurray, B., and Jongman, A. (**2016**). "What comes after /f/? Prediction in speech derives from data-explanatory processes," Psychol. Sci. **27**(1), 43–52.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2002**). "Gradient effects of within-category phonetic variation on lexical access," Cognition **86**(2), B33–B42.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2009**). "Within-category VOT affects recovery from 'lexical' garden paths: Evidence against phoneme-level inhibition," J. Mem. Lang. **60**(1), 65–91.

Meehl, P. E. (**1990**). "Why summaries of research on psychological theories are often uninterpretable," Psychol. Rep. **66**(1), 195–244.

Mermelstein, P. (**1978**). "On the relationship between vowel and consonant identification when cued by the same acoustic information," Percept. Psychophys. **23**, 331–336.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (**2014**). "Phonetic feature encoding in human superior temporal gyrus," Science **343**, 1006–1010.

Meyer, M. K., and Munson, B. (**2021**). "Clinical experience and categorical perception of children's speech," Int. J. Lang. Commun. Disord. **56**(2), 374–388.

Miller, J. L. (**1997**). "Internal structure of phonetic categories," Lang. Cognit. Process. **12**, 865–869.

Miller, J. L., and Eimas, P. D. (**1996**). "Internal structure of voicing categories in early infancy," Percept. Psychophys. **58**(8), 1157–1167.

Miller, J. L., and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category," Percept. Psychophys. **46**(6), 505–512.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**(5), 331–340.

Munson, B. (**2007**). "The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation," Lang. Speech **50**(1), 125–142.

Munson, B., Schellinger, S. K., and Edwards, J. (**2017**). "Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales," Clin. Linguist. Phon. **31**(1), 56–79.

Munson, C. (**2011**). *Perceptual Learning in Speech Reveals Pathways of Processing* (University of Iowa, Iowa City, IA).

Myers, E. B., Blumstein, S. E., Walsh, E., and Eliassen, J. (**2009**). "Inferior frontal regions underlie the perception of phonetic category invariance," Psychol. Sci. **20**(7), 895–903.

Nearey, T. M. (**1997**). "Speech perception as pattern recognition," J. Acoust. Soc. Am. **101**(6), 3241–3254.

Nearey, T. M., and Hogan, J. (**1986**). "Phonological contrast in experimental phonetics: Relating distributions of measurements in production data to perceptual categorization curves," in *Experimental Phonology*, edited by J. Ohala and J. Jaeger (Academic, New York), pp. 141–161.

Newell, F. N., and Bülthoff, H. H. (**2002**). "Categorical perception of familiar objects," Cognition **85**(2), 113–143.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Oden, G., and Massaro, D. W. (**1978**). "Integration of featural information in speech perception," Psychol. Rev. **85**(3), 172–191.

Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**(1), 224–230.

Ou, J., Yu, A. C. L., and Xiang, M. (**2021**). "Individual differences in categorization gradience as predicted by online processing of phonetic cues during spoken word recognition: Evidence from eye movements," Cogn. Sci. **45**(3), e12948.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., and Chang, E. F. (**2012**). "Reconstructing speech from human auditory cortex," PLoS Biol. **10**(1), e1001251.

Phillips, C. (**2001**). "Levels of representation in the electrophysiology of speech perception," Cognit. Sci. **25**(5), 711–731.

Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys. **13**(2), 253–260.

Pisoni, D. B. (**1977**). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," J. Acoust. Soc. Am. **61**(5), 1352–1361.

Pisoni, D. B., and Lazarus, J. H. (**1974**). "Categorical and noncategorical modes of speech perception along the voicing continuum," J. Acoust. Soc. Am. **55**, 328–333.

Pisoni, D. B., and Tash, J. (**1974**). "Reaction times to comparisons within and across phonetic categories," Percept. Psychophys. **15**(2), 285–290.

Plichta, B., and Preston, D. R. (**2005**). "The /ay/s have it the perception of /ay/ as a north-south stereotype in United States English," Acta Linguistica Hafniensia **37**(1), 107–130.

Pollack, I., and Pisoni, D. B. (**1971**). "On the comparison between identification and discrimination tests in speech perception," Psychon. Sci. **24**(6), 299–300.

Quinn, P. C. (**2004**). "Visual perception of orientation is categorical near vertical and continuous near horizontal," Perception **33**(8), 897–906.

Rao, V. N. V., Bye, J. K., and Varma, S. (**2022**). "Categorical perception of *p*-values," Top. Cognit. Sci. **14**, 414–425.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**(1), 81–110.

Rhodes, G., and Leopold, D. A. (**2011**). "Adaptive norm-based coding of face identity," in *Oxford Handbook of Face Perception*, edited by A. Calder, G. Rhodes, M. Johnson, and J. Haxby (Oxford University Press, Oxford, UK), pp. 263-286, available at https://academic.oup.com/edited-volume/28040/chapter-abstract/211939622?redirectedFrom=fulltext.

Roberson, D., Davies, I., and Davidoff, J. (**2000**). "Color categories are not universal: Replications and new evidence from a stone-age culture," J. Exp. Psychol.: General **129**(3), 369–398.

Roberson, D., Hanley, J. R., and Pak, H. (**2009**). "Thresholds for color discrimination in English and Korean speakers," Cognition **112**(3), 482–487.

Robertson, E. K., Joanisse, M. F., Desroches, A. S., and Ng, S. (**2009**). "Categorical speech perception deficits distinguish language and reading impairments in children," Dev. Sci. **12**(5), 753–767.

Rosen, S. M. (**1979**). "Range and frequency effects in consonant categorization," J. Phon. **7**(4), 393–402.

Rosen, S. M., and Howell, P. (**1981**). "Plucks and bows are not categorically perceived," Percept. Psychophys. **30**(2), 156–168.

Sarrett, M., McMurray, B., and Kapnoula, E. (**2020**). "Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and semantic expectations," Brain Lang. **211**, 104875.

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N., and Dupoux, E. (**2021**). "Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input," Proc. Natl. Acad. Sci. U.S.A. **118**(7), e2001844118.

Schouten, M. E. H., Gerrits, E., and Van Hessen, A. (**2003**). "The end of categorical perception as we know it," Speech Commun. **41**, 71–80.

Serniclaes, W., Van Heghe, S., Mousty, P., Carre, R., and Sprenger-Charolle, L. (**2004**). "Allophonic mode of speech perception in dyslexia," J. Exp. Child Psychol. **87**, 336–361.

Slawinski, E. B., and Fitzgerald, L. K. (**1998**). "Perceptual development of the categorization of the /r-w/ contrast in normal children," J. Phon. **26**, 27–43.

Smits, R. (**2001**). "Hierarchical categorization of coarticulated phonemes: A theoretical analysis," Percept. Psychophys. **63**, 1109–1139.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (**1999**). "Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex," J. Neurophysiol. **82**, 2346–2357.

Strand, E. (**1999**). "Uncovering the role of gender stereotypes in speech perception," J. Lang. Social Psychol. **18**, 86–100.

Strange, W., and Shafer, V. L. (**2008**). "Speech perception in second langauge learners: The re-education of selective perception," in

Phonology and Second Language Acquisition, edited by J. G. H. Hansen and M. Zampini (Benjamins, New York), pp. 153–192.

Theodore, R. M., and Monto, N. R. (**2019**). "Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions," Psychon. Bull. Rev. **26**(3), 985–992.

Thiessen, E. D., Girard, S., and Erickson, L. C. (**2016**). "Statistical learning and the critical period: How a continuous learning mechanism can give rise to discontinuous learning," WIREs. Cogn. Sci. **7**(4), 276–288.

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., and Garnsey, S. M. (**2018**). "The time-course of cortical responses to speech revealed by fast optical imaging," Brain Lang. **184**, 32–42.

Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. (**2010**). "Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech," Psychol. Sci. **21**(10), 1532–1540.

Tsuji, S., and Cristia, A. (**2014**). "Perceptual attunement in vowels: A meta-analysis," Dev. Psychobiol. **56**, 179–191.

Van Hessen, A. J., and Schouten, M. E. H. (**1999**). "Categorical perception as a function of stimulus quality," Phonetica **56**, 56–72.

Werker, J. F., and Curtin, S. (**2005**). "PRIMIR: A developmental framework of infant speech processing," Lang. Learn. Develop. **1**(2), 197–234.

Werker, J. F., and Hensch, T. K. (**2015**). "Critical periods in speech perception: New directions," Annu. Rev. Psychol. **66**(1), 173–196.

Werker, J. F., and Tees, R. C. (**1987**). "Speech perception in severely disabled and average reading children," Can. J. Psychol. **41**(1), 48–61.

Whalen, D. H. (**1992**). "Perception of overlapping segments: Thoughts on Nearey's model," J. Phon. **20**, 493–496.

Wyttenbach, R. A., May, M. L., and Hoy, R. R. (**1996**). "Categorical perception of sound frequency by crickets," Science **273**(5281), 1542–1544.

Yeni–Komshian, G. H., and Soli, S. D. (**1981**). "Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation," J. Acoust. Soc. Am. **70**, 966–975.

Zhang, Q., Lei, L., and Gong, T. (**2021**). "Categorical perception as a combination of nature and nurture," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 43.

Zimman, L. (**2017**). "Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants," Linguistics **55**(5), 993–1019.