Journal of Experimental Psychology:
Human Perception and Performance
2017, Vol. 43, No. 9, 1594–1611

0096-1523/17/$12.00 http://dx.doi.org/10.1037/xhp0000410

Evaluating the Sources and Functions of Gradiency in Phoneme Categorization: An Individual Differences Approach

Efthymia C. Kapnoula
University of Iowa

Matthew B. Winn
University of Washington

Eun Jong Kong
Korea Aerospace University

Jan Edwards
University of Wisconsin-Madison

Bob McMurray
University of Iowa

During spoken language comprehension listeners transform continuous acoustic cues into categories (e.g., /b/ and /p/). While long-standing research suggests that phonetic categories are activated in a gradient way, there are also clear individual differences in that more gradient categorization has been linked to various communication impairments such as dyslexia and specific language impairments (Joanisse, Manis, Keating, & Seidenberg, 2000; López-Zamora, Luque, Álvarez, & Cobos, 2012; Serniclaes, Van Heghe, Mousty, Carré, & Sprenger-Charolles, 2004; Werker & Tees, 1987). Crucially, most studies have used 2-alternative forced choice (2AFC) tasks to measure the sharpness of between-category boundaries. Here we propose an alternative paradigm that allows us to measure categorization gradiency in a more direct way. Furthermore, we follow an individual differences approach to (a) link this measure of gradiency to multiple cue integration, (b) explore its relationship to a set of other cognitive processes, and (c) evaluate its role in individuals' ability to perceive speech in noise. Our results provide validation for this new method of assessing phoneme categorization gradiency and offer preliminary insights into how different aspects of speech perception may be linked to each other and to more general cognitive processes.

Public Significance Statement

Labeling sounds and images is an essential part of many cognitive processes that allow us to function efficiently in our everyday lives. One such example is *phoneme categorization*, which refers to listeners' ability to correctly identify speech sounds (e.g., /b/) and is required for understanding spoken language. The present study presents a novel method for studying differences among listeners in how they categorize speech sounds. Our results show that (a) there is substantial variability among individuals in how they categorize speech sounds, and (b) this variability likely reflects fundamental differences in how listeners use the speech signal. The study of such differences will lead to a more comprehensive understanding of both typical and atypical patterns of language processing. Therefore, in addition to its theoretical significance, this study can also help us advance the ways in which we remediate behaviors linked to atypical perception of speech.

Keywords: speech perception, individual differences, categorical perception, multiple cue integration, executive function

Supplemental materials: http://dx.doi.org/10.1037/xhp0000410.supp

This article was published Online First April 13, 2017.

Efthymia C. Kapnoula, Department of Psychological and Brain Sciences, DeLTA Center, University of Iowa; Matthew B. Winn, Department of Speech and Hearing Sciences, University of Washington; Eun Jong Kong, Department of English, Korea Aerospace University; Jan Edwards, Department of Communication Sciences and Disorders, Waisman Center, University of Wisconsin-Madison; Bob McMurray, Department of Psychological and Brain Sciences, Department of Communication Sciences and Disorders, Department of Linguistics, DeLTA Center, University of Iowa.

Jan Edwards is now at the Department of Hearing and Speech Sciences, University of Maryland.

The authors thank Jamie Klein, Dan McEchron, and the students of the MACLab for assistance with data collection. This project was supported by NIH Grant DC008089 awarded to BM.

Correspondence concerning this article should be addressed to Efthymia C. Kapnoula, who is now at the Basque Center on Cognition, Brain and Language, Mikeletegi Pasealekua, 69, 20009, Donostia, Gipuzkoa, Spain. E-mail: kapnoula@gmail.com

Speech varies along multiple acoustic dimensions (e.g., formant frequencies, durations of various events, etc.) that are continuous and highly variable. From this signal, listeners extract linguistically relevant information that serves as the basis of recognizing words. This process represents a transformation from a continuous input that is both ambiguous and redundant into relatively discrete categories, such as features, phonemes, and words.

During this process, listeners face a critical problem: stimuli with the same acoustic cue values[1] may correspond to different categories depending on the context (e.g., speech rate or talker's sex). For example, Voice Onset Time (VOT) is the time between the onset of the release of the articulators and the onset of laryngeal vibration. VOT is the primary cue distinguishing voiced from unvoiced stop consonants with VOTs below 20 ms typical for /b/,/d/,/g/, while VOTs over 20 ms are typical for /p/,/t/,/k/. However, contextual factors can make VOT more ambiguous. For example, a stimulus with a VOT of 20 ms could be a /b/ in slow speech or a /p/ in fast speech. Despite over 40 years of research, speech scientists have identified few (if any) acoustic cues that unambiguously identify a phoneme across different contexts (e.g., McMurray & Jongman, 2015; Ohala, 1996).

Traditional approaches suggest that this problem is solved via specialized mechanisms that discard irrelevant information, leading to the perception of distinct phonemic categories (Liberman, Harris, Hoffman, & Griffith, 1957). However, recent studies show that typical listeners maintain information that is seemingly irrelevant for discriminating between phonemic categories (i.e., within-category information; Massaro & Cohen, 1983; McMurray, Tanenhaus, & Aslin, 2002; Toscano, McMurray, Dennhardt, & Luck, 2010).

Recent theoretical approaches suggest that such gradient representations may be useful for coping with ambiguity and integrating different pieces of information (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015; McMurray & Farris-Trimble, 2012; Oden & Massaro, 1978). However, there is little empirical data that speaks to the functional role of maintaining within-category information (though see McMurray, Tanenhaus, & Aslin, 2009).

The present study addresses this issue using an individual differences approach. Work by Kong and Edwards (2011, 2016) suggests listeners vary in the degree to which they maintain within-category information (i.e., how gradiently they categorize speech sounds). We examined these individual differences and their role in speech perception by (a) linking them to a different aspect of speech perception (the use of secondary acoustic cues), (b) investigating their potential sources in executive function, and (c) examining how they relate to speech perception in noise.

The Problem of Lack of Invariance and Categorical Perception

Variability in the acoustic signal is commonly described in terms of acoustic/phonetic cues such as VOT. Critically, while acoustic cues are *continuous*, our percept (as well as most linguistic analyses) reflects more or less *discrete* categories (/b/ and /p/). Mapping continuous cues onto discrete categories is complex because the same cue values can map onto different categories, depending on many factors, including the talker's gender (Hillenbrand, Getty, Clark, & Wheeler, 1995), neighboring speech sounds (coarticulation, Hillenbrand, Clark, & Nearey, 2001), and speaking rate (Miller, Green, & Reeves, 1986). This is the problem of *lack of invariance*: speech sounds do not have invariant acoustic attributes, and a single acoustic cue cannot be reliably mapped to a single speech sound.

One solution to the lack of invariance problem was suggested by *Categorical Perception* (CP; Liberman et al., 1957). CP describes the well-established behavioral phenomenon that discrimination within a category (e.g., between two instances of a /b/) is poor, but discrimination of an equivalent acoustic difference that spans a category boundary is quite good (e.g., Liberman Harris, Kinney, & Lane, 1961; Pisoni & Tash, 1974; Schouten & van Hessen, 1992; Repp, 1984 for a review; and see Chang et al., 2010; Dehaene-Lambertz, 1997; Phillips et al., 2000; Sams, Aulanko, Aaltonen, & Näätänen, 1990, for related neural evidence).

One hypothesis is that CP derives from some form of warping of the perceptual space that amplifies the influence of categories. Under this view, a [b] with a VOT of 15 ms is encoded more similarly to one with a VOT of 0 ms than to a [p] with a VOT of 30 ms. Such warping is often attributed to specialized processes that discard within-category variation in favor of discrete encoding at both the auditory/cue level and at the level of phoneme categories. This view—perhaps best exemplified by *motor theory* (Liberman & Whalen, 2000)—suggests that auditory encoding is aligned to the discrete goals of the system (phoneme categorization). As a result, acoustic variation, arising from talker differences and/or coarticulation, does not pose a challenge for speech perception, because the *underlying* representations (gestures or phonological units) can be rapidly extracted by such specialized mechanisms.

The Gradient Alternative

According to CP, encoding of acoustic cues is somewhat discrete, and, this enables cues to be easily mapped to fairly discrete categories. However, the claim of discreteness at both levels has not held up to scrutiny. Serious concerns have been raised about the discrimination tasks used to establish CP, while the degree to which discrimination is categorical (i.e., better discrimination across a boundary) depends on the specific task (Carney, Widin, & Viemeister, 1977; Gerrits & Schouten, 2004; Pisoni & Lazarus, 1974; Schouten, Gerrits, & van Hessen, 2003). Gerrits and Schouten (2004) and Schouten et al. (2003) found that working memory demands associated different tasks can lead listeners to rely on subjective labels (rather than auditory codes, which may decay more rapidly). A reliance on labels could lead to a more categorical pattern of responses, even if the precategorized perceptual representation is continuous (see also Carney et al., 1977; Gerrits & Schouten, 2004; Pisoni & Tash, 1974). Therefore, CP may in fact reflect the influence of categories on memory and decision processes, not on perceptual processes per se. Indeed, when less biased discrimination measures are employed, CP-like

[1] Even though we use the term "cues" here, we do not make a strong theoretical commitment as to the kind of auditory information this term entails.

effects disappear (Gerrits & Schouten, 2004; Massaro & Cohen, 1983; Pisoni & Lazarus, 1974).

This dependence of CP on the discrimination task implies that encoding of speech cues may not be warped at all, but rather may reflect the input monotonically. Consistent with this idea, ERP and MEG responses to isolated words from VOT continua reflect a linear pattern of response to changes along the continuum with no evidence of warping (Frye et al., 2007; Toscano et al., 2010; and see Myers, Blumstein, Walsh, & Eliassen, 2009 for MRI evidence). Moreover, beyond auditory encoding, there is substantial evidence that fine-grained detail is preserved at higher levels of the pathway, affecting even lexical processing (Andruski, Blumstein, & Burton, 1994; McMurray et al., 2002; Utman, Blumstein, & Burton, 2000).

The Functional Role of Gradiency

The usefulness of maintaining within-category information throughout levels of processing is a key idea of several theoretical approaches (Goldinger, 1998; Kleinschmidt & Jaeger, 2015; Kronrod, Coppess, & Feldman, 2016; McMurray & Jongman, 2011; Oden & Massaro, 1978). It is hypothesized to allow for more flexible and efficient speech processing via at least three mechanisms. First, processes such as coarticulation and assimilation leave fine-grained, subcategorical traces in the signal (e.g., Gow, 2001), which can be used to anticipate upcoming input, speeding up processing. Multiple studies suggest that listeners take advantage of anticipatory coarticulatory information in this way (Gow, 2001; Mahr, McMillan, Saffran, Ellis Weismer, & Edwards, 2015; McMurray & Jongman, 2015; Salverda, Kleinschmidt, & Tanenhaus, 2014; Yeni-Komshian & Soli, 1981). As these modifications are largely within-category, such anticipation is only possible if listeners are sensitive to this fine-grained detail.

Second, gradient encoding may offer greater flexibility in how cues map onto categories (e.g., Massaro & Cohen, 1983; Toscano et al., 2010). Continuous encoding of cues may, for example, permit for the values of one cue (e.g., VOT) to be interpreted in light of the values of another cue (e.g., F_0). Such processes may underlie the well-known trading relations that have been documented in speech perception (Repp, 1982; Summerfield & Haggard, 1977; Winn, Chatterjee, & Idsardi, 2013). This kind of combinatory process would also be necessary for accurately compensating for higher level contextual expectations—for example, recoding pitch relative to the talker's mean pitch (McMurray & Jongman, 2011, 2015).

Third, gradient responding at higher levels, at the level of phonemes (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; Miller, 1997); or words (Andruski et al., 1994; McMurray et al., 2008) may help cope with uncertainty. With a gradient encoding, the degree to which the perceptual system commits to one representation over another (e.g., /b/ vs. /p/) is monotonically related to the degree of support in the signal. For example, a labial stop with a VOT of 5 ms activates /b/-onset words *more* than a labial stop with a VOT of 15 ms, even though both map onto the same category. Superficially, this may appear disadvantageous as it could slow an efficient decision. However, given the variability, and noise in the signal, gradiency may allow listeners to "hedge" their bets in the face of ambiguity. It is precisely when cue values are more ambiguous that listeners should not commit too strongly and keep their options open until more information arrives (Clayards et al., 2008; McMurray et al., 2009).

In sum, gradiency may allow the system to (a) harness fine-grained (within-category) differences that may be helpful; (b) integrate information from multiple sources more flexibly; and (c) delay commitment when insufficient information is available. Thus, while the somewhat empirical question of the gradient versus discrete nature of speech representations has been hotly debated (Chang et al., 2010; Gerrits & Schouten, 2004; Liberman & Whalen, 2000; Massaro & Cohen, 1983; McMurray et al., 2002; Myers et al., 2009; Toscano et al., 2010), it has important theoretical ramifications for how listeners solve a fundamental perceptual problem.

Individual Differences in Phoneme Categorization

Despite the evidence for gradiency in typical listeners, it is less clear whether there are individual differences. Mounting evidence now exists in neuroscience for multiple pathways of speech processing (Blumstein, Myers, & Rissman, 2005; Hickok & Poeppel, 2007; Myers et al., 2009) that can be flexibly deployed under different conditions (Du, Buchsbaum, Grady, & Alain, 2014). Given this, different listeners may adopt different solutions to this problem, perhaps providing more weight to either dorsal or ventral pathways (see Ojemann, Ojemann, Lettich, & Berger, 1989 for analogous evidence in word production). Similarly, the Pisoni and Tash (1974) model of CP suggests that listeners have simultaneous access to both continuous acoustic cues *and* discrete categories. Again, this raises the possibility that listeners may weight these two sources of information differently during speech perception.

Considering the function of gradiency in speech perception, the possibility of individual differences raises three questions: (a) Are listeners gradient to varying degrees? (b) What are the sources of these differences? (c) Do such differences impact speech perception as a whole?

Much of the debate around categorical versus gradient perception in typical listeners concerns the degree to which gradiency might be adaptive (or maladaptive). In this regard, a consideration of listeners with communication disorders may be useful. Work on language-related disorders such as specific language impairment (SLI) and dyslexia suggests significant differences in the gradiency of speech perception between impaired and typical listeners (Coady, Evans, Mainela-Arnold, & Kluender, 2007; Robertson, Joanisse, Desroches, & Ng, 2009; Serniclaes, 2006; Sussman, 1993; Werker & Tees, 1987, but see Coady, Kluender, & Evans, 2005; McMurray, Munson, & Tomblin, 2014). Much of this work has examined phoneme categorization in a 2-alternative forced choice (2AFC) task. In this task, participants hear a word (or phoneme sequence; e.g., *ba* or *pa*) from a continuum ranging in small steps from one endpoint to the other and assign one of two labels. Listeners typically show a sigmoidal response function with a sharp transitioning from one phoneme category to the other. Critically, the steepness of the slope of the response function is used as a measure of category discreteness.

Using this measure, impaired listeners generally show shallower transitions between categories (but see Blomert & Mitterer, 2004; Coady, Kluender, & Evans, 2005; McMurray et al., 2014). For example, Werker and Tees (1987) found that children with reading difficulties had shallower slopes on a /b/-to-/d/ continuum than typical children (see also Godfrey, Syrdal-Lasky, Millay, & Knox, 1981; Serniclaes, Sprenger-Charolles, Carré, & Demonet, 2001). Joanisse, Manis, Keating, and Seidenberg (2000) found a similar pattern for language impaired (LI) children. More recently, López-Zamora, Luque, Álvarez, and Cobos (2012) found that shallower slopes in a phoneme identification task predict atypical syllable frequency effects in visual word recognition, suggesting some kind of atypical pattern of sublexical processing. Lastly, Serniclaes, Ventura, Morais, and Kolinski (2005) found that illiterate adults have shallower identification slopes than literate ones.

These findings are typically attributed to nonoptimal CP; if impaired learners encode cues inaccurately (e.g., they hear a VOT of 10 ms occasionally as 5 or 15 ms), then tokens near the boundary are likely to be encoded with cue values on the other side, flattening the function. This assumes a sharp, discrete category boundary as the optimum response function, which is corrupted by internal noise (in the encoding of acoustic cues) for disordered listeners (Moberly, Lowenstein, & Nittrouer, 2016; Winn & Litovsky, 2015). Thus, impaired listeners may have equally sharp underlying categorization functions as nonimpaired listeners, but the categorization output is corrupted due to noisier auditory encoding.

This account offers a clear explanation for listeners with obvious sensory impairments (e.g., hearing impairment), however, it may be less compelling, in the case of listeners with dyslexia or SLI, who may have impairments at higher levels than cue encoding. One alternative explanation is that children with dyslexia have *heightened* within-category discrimination (Werker & Tees, 1987). This links dyslexia to a difficulty in discarding acoustic detail that is linguistically irrelevant (Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008; Serniclaes et al., 2004), a failure of a functional goal of categorization. Even in this case, however, the assumption is that discrete categorization, and a reduction of within-category sensitivity are to be desired, and a failure of any aspect of this process drives the shallower response slope (but see Messaoud-Galusi, Hazan, & Rosen, 2011).

Few studies have examined individual differences from the perspective that gradient perception may be beneficial (though see McMurray et al., 2014). An exception is Clayards et al., (2008) who manipulated within-category variability of VOT across trials. When VOTs were more variable, listeners' response patterns followed shallower 2AFC slopes. This suggests that a shallower identification slope may reflect a different (and more useful) way of mapping cue values onto phoneme categories in that it reflects uncertainty in the input.

It is not clear how to reconcile the classic (categorical) view, arguing for the utility of more categorical labeling functions, with the more recent view that gradiency may be beneficial. Both sides may hold truth; shallower functions could derive from both noisier cue encoding *and* a more graded mapping of cues to categories. What is clear from the work on disordered language is that group differences in categorization relate to differences in language processing. More importantly, our review suggests that measures such as 2AFC phoneme identification may not do a good job measuring these differences, because it is difficult to distinguish noisy cue encoding from more gradient categorization.

Toward a New Measure of Phoneme Perception Gradiency

The foregoing review reveals a fundamental limitation of 2AFC tasks: the systematicity with which listeners identify acoustic cues and map them to phoneme categories (noise) may be orthogonal to the degree to which they maintain within-category information (López-Zamora et al., 2012; Messaoud-Galusi et al., 2011). This is partly because the 2AFC task only allows binary responses. When a listener reports a stimulus as /b/ 30% of the time and as /p/ 70%, it could be because they discretely thought the stimulus was a /b/ on 30% of trials, or because they thought it had some likelihood of being either or both (on every trial) and the responses reflect the probability distributions of cues-to-categories mappings. A continuous measure may be more precise; if listeners hear the stimulus categorically as /b/ on 30% of trials, the trial-by-trial data should reflect a fully /b/-like response on those trials. In contrast, if listeners' representations reflect partial ambiguity, they should respond in between with variance clustered around the mean rating. As Massaro and Cohen (1983) argue: "relative to discrete judgment, continuous judgments may provide a more direct measure of the listener's perceptual experience".

One such task is a visual analogue scaling (VAS) task. In this task, participants hear an auditory stimulus and select a point on a line to indicate how close the stimulus was to the word on each side (Figure 1; Massaro & Cohen, 1983, for an analogous task in discrimination). This continuous response (instead of a binary choice) permits a more direct measure of gradiency. For example, if we assume a step-like categorization function plus noise in the cue encoding, listeners' responses should cluster close to the extremes of the scale, though for stimuli near the boundary, participants might choose the *wrong* extreme because noise would lead to misclassifications (e.g., they may choose the left end of the continuum for ambiguous /p/-initial stimuli). On the other hand, if listeners respond gradiently, we should observe a more *linear* relationship between the cue value (e.g., the VOT) and the VAS response, with participants using the whole range of the line and variance across trials clustering around the line. However, under either model, a 2AFC would give us an identical response function: a shallower slope.

VAS tasks have been used in speech, generally supporting the gradient perspective. Massaro and Cohen (1983) used a VAS task to show that discrimination continuously related to acoustic distance without warping by categories. Many studies by Miller and colleagues (e.g., Allen & Miller, 1999; Miller & Volaitis, 1989) used a VAS goodness scale task (e.g., asking "*How good of a /p/ was this?*") to characterize the graded prototype structure of phonetic categories. However, none of these lines of work examined individual differences or related such measures to variation in 2AFC categorization.

Kong and Edwards (2011, 2016), building on related work by Schellinger, Edwards, Munson, and Beckman (2008) and Urberg-Carlson, Kaiser, and Munson (2008), offer evidence for individual

Figure 1. Visual analogue scaling task used by Kong and Edwards (2011, 2016).

differences (see also Schellinger, Munson, & Edwards, 2016). They tested adults on a /da/-/ta/ continuum, asking them to rate each token on a continuous scale. Participants varied in their ratings; some exhibited a more categorical pattern, preferring the endpoints of the line, while others were more gradient, using the entire scale. Further, more gradient responders showed a stronger reliance on a secondary acoustic cue in a separate categorization task—a pattern that was consistent across two separate testing sessions. Lastly, there was a correlation between gradiency and cognitive flexibility (assessed by the switch version of the Trail Making task), suggesting a link between speech perception and executive function.

These findings speak to the potential strengths of an individual differences approach for studying fundamental aspects of speech perception. Kong and Edwards (2011, 2016) demonstrate the reliability of VAS measures, and provide preliminary support for a link between gradiency and the use of secondary cues (a key prediction of accounts suggesting gradiency could be beneficial to speech perception). However, some important methodological refinements and experimental extensions are necessary to fully address the key questions we ask here.

First, to assess secondary cue use, Kong and Edwards used the anticipatory eye movement (AEM) task (McMurray & Aslin, 2004). This is a somewhat nontraditional measure of phoneme categorization that makes it difficult to evaluate their results in relation to studies using more traditional (e.g., 2AFC) measures of phoneme categorization. It is, therefore, unclear how the same individual may perform the more traditional 2AFC task versus a task such as the VAS, and the differences between the two patterns of performance would inform our understanding of the speech perception processes these two tasks tap into.

The previous point is particularly important given the discrepancy between studies of language disorders that have found shallower 2AFC slopes (e.g., Werker & Tees, 1987), and the newer view from basic research showing that gradiency is the typical pattern in nonimpaired listeners and may be adaptive (Clayards et al., 2008). The VAS task may offer unique insight into the relationship between the 2AFC task and these contrasting theoretical views of gradiency.

The second motivation for the current study is arguably the most important; Kong and Edwards's (2011) statistical measure of gradiency captured the overall distributions of ratings (e.g., how often participants use the VAS endpoints) independently of the stimulus characteristics. While this documents individual differences, it may also be limited for two reasons. First, it leaves open the possibility that individual differences may also be sensitive to other aspects of speech perception (e.g., multiple cue integration or noise). For example, a flatter distribution could be obtained if listeners matched their VAS ratings to the VOT, or if they showed a large effect of F_0 (which would spread out their responses), or even if they simply guessed. In contrast, by taking into account the stimulus (e.g., the VOT) we can estimate categorization gradiency independently of potentially confounding factors. Second, by developing a stimulus-dependent measure we can also compute an estimate of trial-by-trial noise in the encoding of stimuli, addressing a main critique of the 2AFC task.

Finally, executive function (EF) is a multifaceted construct. Kong and Edwards used two measures (Trail Making and color-word Stroop), which possibly load on different aspects of EF, but only found a correlation between the former and gradiency (though this should be qualified by their moderate sample size of 30). One goal here was to employ additional measures of EF, along with a much larger sample size to obtain a more definitive answer to this question.

Thus, the present study built on the Kong and Edwards VAS paradigm, but addressed the aforementioned issues with a number of changes and refinements of the methodology, including the use of a novel technique specifically developed to help us disentangle categorization gradiency from other aspects of speech perception.

The Present Study

We sought to examine individual differences in speech perception by (a) establishing a precise and theoretically grounded measure of gradiency from the VAS task, (b) exploring the role of several factors that may be linked to these differences, and (c) assessing the role of gradiency in the perception of speech in noise (an issue not addressed by prior studies).

We collected VAS responses from a large sample ($N = 131$), so that we could better evaluate individual differences in phoneme categorization gradiency. Listeners heard tokens from a two-dimensional voicing continuum (matrix) that simultaneously varied in VOT and F_0 (a secondary cue) and rated each token (how *b*-like vs. *p*-like it sounded) using the VAS. Critically, we developed and validated a new set of statistical tools for assessing an individual subject's gradiency that captured gradiency in responding in the same model that captured the relationship between stimulus-related factors and VAS responses.

Secondarily, we used a variety of continua (word and nonword, labial- and alveolar-initial) to assess the effects of lexical status and place of articulation respectively. While these manipulations were exploratory, prior results suggest that listeners may be more sensitive to subphonemic detail in real words (McMurray et al., 2008). This raises the possibility that the individual differences reported by Kong and Edwards are only seen with nonwords, while most listeners show a gradient response pattern with words.

Next, we related our gradiency measure to the more standard 2AFC measure of categorization. As described, the 2AFC slope may reflect both categorization gradiency and internal noise in cue encoding. Thus, an explicit comparison between the VAS and 2AFC tasks may help disentangle what the 2AFC task is primarily measuring. Since both tasks are thought to reflect, at least to some degree, categorization gradiency, we expected a positive correlation between the VAS and 2AFC slopes. However, it was not clear how strong a correlation should be expected, given the ambiguity as to what affects the 2AFC task.

We also related gradiency (in the VAS task) to cue integration (from the 2AFC task), indexed by the influence of a secondary cue on categorization. As described above, we predicted that gradient listeners would be more sensitive to fine-grained information and should, therefore, be better at taking advantage of multiple cues (see Kong & Edwards, 2016).

Next, we extended earlier investigations by addressing whether these speech measures (gradiency and multiple cue integration) were related to nonlinguistic cognitive abilities. We collected a set of individual differences measures tapping different aspects of executive function to evaluate these higher cognitive processes as possible (direct or indirect) sources of gradiency. Our hypothesis was that, to the extent that speech categorization may draw on domain-general skills such as EF or working memory, individual differences in these skills may be reflected in the gradiency or discreteness of categorization.

Finally, we performed a preliminary assessment of the functional role of gradiency (i.e., whether it is beneficial for speech perception) using a speech-in-noise recognition task.

Method

Participants

Participants were 131 adult monolingual speakers of American English, all of whom completed a hearing screening at four octave-spaced audiometric test frequencies for each ear; one participant was excluded on this basis because of thresholds greater than 25 dB HL. Participants received course credit, and underwent informed consent in accord with University of Iowa IRB policies. Technical problems with several tasks led to their results not being available for one or more participants. Consequently, between two and 11 participants were excluded from the analyses of the specific tasks for which there were missing data.

Overview of Design

Participants performed six tasks (see Table 1). To explore stimulus-driven effects on gradiency, we included voicing continua for labials and alveolars (within subject) in words, nonwords, and phonotactically impermissible nonwords (between subjects). VAS stimuli varied on seven VOT steps and five F_0 steps (secondary cue).

A conventional 2AFC task was compared to the more continuous VAS task. The VAS task was always performed before the 2AFC task to avoid inducing any step-like bias on the former by the latter. The 2AFC task was conducted on continua that varied on seven steps of VOT and only two steps of F_0; this allowed an independent estimate of secondary cue use measured as the difference in the category boundary between the two VOT continua.

We used three measures of nonlanguage cognitive function, tapping different aspects of executive function (EF). We used the Flanker task to assess inhibition, the N-Back task, which taps primarily working memory, and the Trail Making task as a measure of planning and executive performance. Finally, as a measure of speech perception accuracy, we administered a computerized version of the AzBio sentences (Spahr et al., 2012), a speech-in-noise measure.

Measuring Phoneme Categorization Gradiency

To measure individual differences in phoneme categorization gradiency we used the VAS task with three types of continua (*stimulus-types*): (a) consonant-vowel-consonant (CVC) real words (RW); (b) CVC nonwords, (NW); and (c) phonotactically impermissible nonword CVs[2] that violated an English phonotactic constraint that lax vowels cannot appear in open syllables. Each participant was only tested on one stimulus-type (randomly selected). Within that, each participant was tested on two places of articulation (PoA), labial (e.g., *bull-pull*) and alveolar (e.g., *ten-den*; see Table 2).

VAS stimuli and design. For each of the six pairs (see Table 2) we created a two-dimensional continuum by orthogonally manipulating VOT and F_0 *in Praat (*Boersma & Weenink, 2016; [version 5.3.23]). VOT were manipulated in natural speech using progressive cross-splicing (Andruski et al., 1994; McMurray et al., 2008). Progressively longer portions of the onset of a voiced sound (/b/ or /d/) were replaced with equivalent amounts from the aspiration of the voiceless sound (/p/ or /t/). Prior to splicing, voicoids were multiplied by a 3 ms onset ramp, and cross-spliced with the consonant burst/aspiration segment using a symmetrical 2-ms cross-fading envelope, to remove any waveform discontinuities at the splice point.

At each VOT step, the pitch contour was extracted and modified using the pitch-synchronous overlap-add (PSOLA) algorithm in Praat. Pitch onset varied in five steps spaced equally from 190 Hz to 125 Hz. Pitch was kept steady over the first two pitch periods of the vowel and fell (or rose) linearly until returning to the original contour 80 ms into the vowel. Final stimuli varied along seven VOT steps (1 to 45 ms) and five F_0 steps (90 to 125 Hz). During the VAS task, each participant was presented with all 35 stimuli from each of the two PoA series with three repetitions of each stimulus resulting in 210 trials. Stimulus presentation was blocked by PoA, with the block order counterbalanced between participants.

VAS procedure. On each trial, participants saw a line with a printed word at each end (e.g., *bull* and *pull*, Figure 1). Voiced-initial stimuli were always on the left side. Participants used a computer mouse to click on a vertical bar and drag it from the center to a point on the line to indicate where they thought the sound fell in between the two words. Before starting, participants performed a few practice trials. Unless the participant had clarifying questions, no further instructions were given. The VAS task took approximately 15 min.

Preprocessing of VAS data. One obvious analytic strategy would be to fit a logistic to each participant's VAS data and use the slope as a measure of gradiency. However, since stimuli also varied along a secondary cue, this method is problematic; if a listener has a discrete boundary in VOT space, but the location of this boundary varies with F_0, the average boundary (across F_0s) would look gradient. Instead what is needed is a two-dimensional estimate of the slope. While logistic regression can handle this by weighting and summing the two independent factors, there is no single term separating the contribution of each cue from the overall slope.

To solve this problem, we developed a new function (Eq.1), the *rotated logistic*. This assumes a diagonal boundary in a two-dimensional space described by a line with some cross-over point (along the primary cue) and an angle, θ (see Figure 2). A 90° θ indicates exclusive use of the primary cue, while a 45° θ indicates roughly equal use of both cues. Once θ is identified, we rotate the coordinate space to be orthogonal to this boundary and estimate the slope of the response function perpendicular to the boundary.

This allows us to model gradiency with a single parameter that reflects the derivative of the function orthogonal to the diagonal boundary; shallower slopes indicate more gradient responses, independently of cue use (see Figure 3).

[2] Similar to those used by Kong and Edwards (2011, 2016)

Table 1
Order and Description of Tasks

Order	Task	Domain	Primarily measure of . . .
1	VAS	Speech categorization	Phoneme categorization gradiency
2	Flanker	Cognitive	Executive function: inhibitory control
3	N-Back	Cognitive	Executive function: working memory
4	2AFC	Speech categorization	Secondary cue use
5	Trail Making	Cognitive	Executive function: general
6	AzBio	Speech perception in noise	Speech perception accuracy

Note. VAS = visual analogue scaling; 2AFC = 2-alternative forced choice.

$$p(resp) = b_1 + \frac{(b_2 - b_1)}{1 + e^{\left(\frac{-4 \cdot s \cdot 2 \cdot \upsilon(\theta)}{(b_2 - b_1)}\right) \cdot \left(\frac{\tan(\theta) \cdot (x_0 - VOT) - F_0)}{\sqrt{1 + \tan(\theta)2}}\right)}} \tag{1}$$

Here, b_1 and b_2 are the lower and upper asymptotes, and s is the slope (as in the four-parameter logistic). The new parameters are: θ, the angle of the boundary (in radians), and x_o, is the boundary's x-intercept. The independent variables are VOT and F_o. $\upsilon(\theta)$ (in the denominator, seen in [2]) switches the slope direction if θ is less than 90° to keep the function continuous.

$$\upsilon(\theta) = \begin{cases} 1 & \text{if } \theta < = (\pi/2) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

For each participant, we calculated the average of the three responses for each of the 70 stimuli participants heard during the VAS task (separately for each PoA). The equation in (1, 2) was then fit to each subject's averaged VAS data using a constrained gradient descent method implemented in Matlab (using FMINCON) that minimized the least squared error (see S.1 for details about the curvetting procedure).

To assess the validity of this procedure, we conducted extensive Monte Carlo analyses. These tested both the ability of this procedure to estimate the true values of the data, and looked for any spurious correlations imposed on the data by the function or the curve fitting (e.g., if parameters were confounded with each other). These are reported in supplement S.2 and show very high validity, and no evidence of spurious correlation between the estimated parameters.

Measuring Multiple Cue Integration

We used a 2AFC task for two purposes. First, it offered a measure of multiple cue integration that is independent of the VAS. Second, by relating VAS slope to categorization slope we hoped to determine what drives changes in categorization slope.

Table 2
Stimuli Used in the VAS and the 2AFC Tasks

Place of articulation of first phoneme	Stimulus type		
	Real word	Nonword	CV
Labial	bull—pull	buv—puv	buh—puh
Alveolar	den—ten	dev—tev	deh—teh

Note. VAS = visual analogue scaling; 2AFC = 2-alternative forced choice; CV = consonant-vowel.

2AFC stimuli and design. The 2AFC task was performed immediately after the N-Back task for all participants. A subset of the VAS stimuli was used in the 2AFC task: all 7 VOT steps, but only the two extreme F_0 values. This simplified quantification of listeners' use of F_0 as the difference between boundaries for each F_0. Each of the 28 (7 VOTs × 2 F_0s × 2 PoA) stimuli was presented 10 times (280 total trials). Similarly to the VAS task, trials were presented in two separate blocks, one for each PoA, and block order was counterbalanced between participants.

2AFC procedure. On each trial participants saw two squares, one on each side of the screen, each containing one of two printed words (e.g., *bull/pull*). The voiced-initial word was always in the left square. Participants were prompted to listen carefully to each stimulus and click in the box with the word that best matched what they heard. At the beginning of the task participants performed a few practice trials. The 2AFC task took approximately 11 min.

Preprocessing of 2AFC data. To quantify F_0 use, we fitted each participant's response curve using a four-parameter logistic function (see McMurray, Samelson, Lee, & Bruce Tomblin, 2010) that provides estimates for minimum and maximum asymptotes, slope, and crossover (see Eq. 3).

$$p(resp) = b_1 + \frac{b_2 - b_1}{1 + e^{\left(\frac{-4 \cdot s}{(b2 - b1)}(x - co)\right)}} \tag{3}$$

In this equation, b_1 is the lower asymptote, b_2 is the upper asymptote, s is the slope, and co is the x-intercept (see hypothetical data in Figure 4). This function was fit to each participant's responses separately for each F_0 and for each PoA. Curves were fit using a constrained gradient descent method implemented with FMINCON in Matlab.

Measures of Executive Function

To investigate whether individual differences in cognitive function are related to gradiency in phoneme categorization, we used three tasks measuring aspects of executive function: (a) the Flanker task (available through the NIH Toolbox; Gershon et al., 2013), (b) the N-Back task, and (c) the switch version of the Trail Making task (TMT-B).

Flanker task (inhibitory control). The Flanker task is commonly considered a measure of inhibitory control (Eriksen & Eriksen, 1974). Participants saw five arrows at the center of the screen and reported the direction of the middle arrow by pressing a key. The direction of the other four surrounding arrows (flankers) was either consistent or inconsistent with the target. On inconsis-

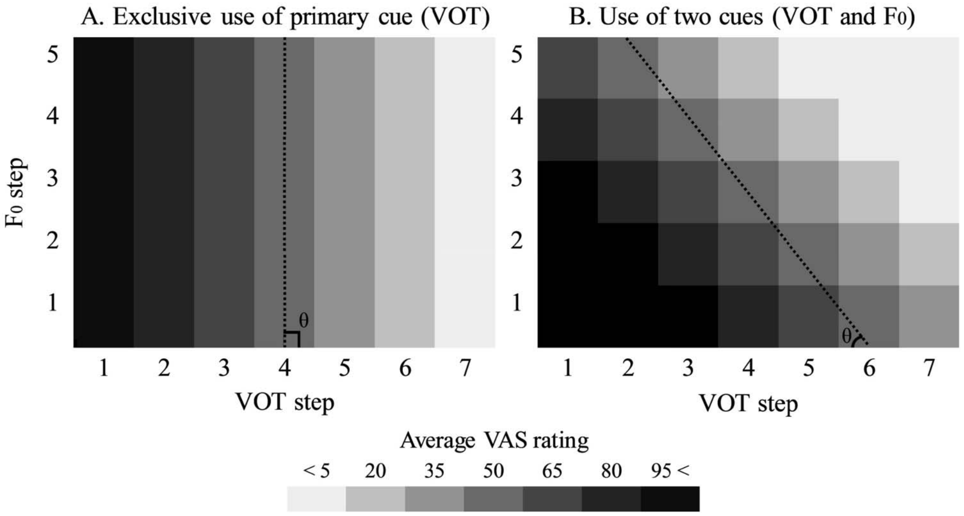


Figure 2. Hypothetical response patterns based on mono-dimensional (left) and bidimensional (right) category boundaries. VOT = Voice Onset Time; VAS = visual analogue scaling.

tent trials, the degree to which participants inhibit the flanking stimuli predicts response speed. The Flanker task had 20 trials (approximately 3 min). Inhibition measures were a composite of both speed and accuracy, following NIH toolbox guidelines.[3]

N-Back task (working memory). The N-Back task was used to measure complex working memory (Kirchner, 1958). Participants viewed a series of numbers (each presented for 2,000 ms) and indicated whether the current number matched the previous one (1-back), the one two numbers before (2-back), or three before (3-back). The three levels of difficulty were presented in this order for all participants. There were 41, 42, and 43 trials for each difficulty level (respectively), yielding 40 responses to be scored in each level. The N-Back task took approximately 9 min. Average accuracy across the three difficulty levels was used as an indicator of working memory capacity.

Trail Making task (cognitive control). Part B of the Trail Making task assesses cognitive control (Tombaugh, 2004). During this task participants were given a sheet of paper with circles containing numbers 1 through 16 and letters A through P. They used a pencil to connect the circles in order, alternating between numbers and letters, starting at number 1 and ending at letter P. The time it took to complete this task was recorded by a trained examiner and used as a measure of cognitive control. On average, the Trail Making task took 2.5 mins to administer.

Speech Recognition in Noise: The AzBio Sentences

To measure how well participants perceive speech in noise, we administered the AzBio sentences (Spahr et al., 2012), which consists of 10 sentences masked with multitalker babble (0 dB SNR). Sentences were delivered over high-quality headphones and participants repeated each sentence with no time constraint. An examiner recorded the number of correctly identified words on a computer display by clicking on each word of the sentence that was correctly produced. The logit-transformed percentage of correctly identified words was used as a measure of overall performance. The AzBio task took approximately 7 min.

Results

We start with a brief descriptive overview of the VAS and the 2AFC data to validate the tasks and examine stimulus factors such as the role of word/nonword status. We then proceed to our theoretical questions.

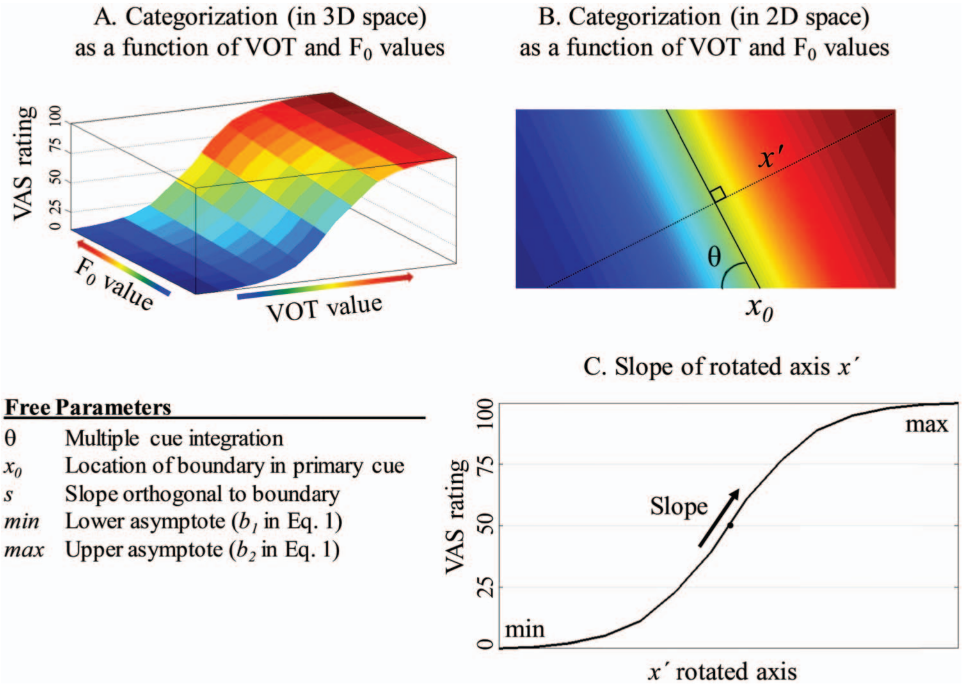
Descriptive Overview of VAS Data

Participants performed the VAS task as instructed, except three who responded with random points on the line and were excluded from analyses. In addition, technical problems led to missing data for five participants, leaving 123 participants with data for this task.

Participants used both VOT and F_0 to categorize stimuli. As expected, participants rated stimuli with higher VOT and higher F_0 values as more /p/ (or /t/) like (see Figure 5). Replicating Kong and Edwards (2011), participants differed substantially in how they performed the VAS task. This can be clearly seen by computing simple histograms of the points that were used along the scale. As Figure 6A shows, some participants primarily responded using the endpoints of the line (Figure 6A), suggesting a more categorical mode of responding, while others used the entire line (Figure 6B), suggesting a more gradient pattern.

While histograms such as those shown in Figure 6 show individual differences, this approach cannot address our primary questions because it ignores the stimulus. For example, Subject 9 could show a flat distribution because they guessed or because they aligned VAS ratings with the stimulus characteristics. A better approach must consider the relationship between stimulus and response.

[3] Flanker task accuracy score = 0.125 * Number of Correct Responses; Reaction Time (reaction time [RT]) Score = 5-(5*[(log[RT]-log(500)]/(log(3000-log[500]). If accuracy levels are ≤ 80%, the final “total” computed score is the accuracy score. If accuracy levels are >80%, RT score and accuracy score are combined.


A. Categorization (in 3D space) as a function of VOT and F0 values
VAS rating
0 25 50 75 100
F0 value
VOT value
B. Categorization (in 2D space) as a function of VOT and F0 values
x′
θ
x0
C. Slope of rotated axis x′
VAS rating
0 25 50 75 100
max
Slope
min
x′ rotated axis

Free Parameters

θ	Multiple cue integration
x_0	Location of boundary in primary cue
s	Slope orthogonal to boundary
min	Lower asymptote (b_1 in Eq. 1)
max	Upper asymptote (b_2 in Eq. 1)

Figure 3. Measuring phoneme categorization gradiency using the *rotated logistic*; Panel A: 3D depiction of voiced/unvoiced stop categorization as a function of VOT and F_0 information (blue/lower front edge at more voiced VAS rating; red/high back edge at more unvoiced VAS rating); Panel B: 2D depiction of the same categorization function; θ marks the theta angle that we use to rotate the x-axis so that it is orthogonal to the categorization boundary; Panel C: depiction of categorization slope using the rotated x-axis (x'). VOT = Voice Onset Time; VAS = visual analogue scaling. See the online article for the color version of this figure.

Figure 7 shows results for two participants plotting the individual (trial by trial) VAS responses as a function of VOT and F_0. Subject 7 gives mostly binary responses, VAS scores near 0 or 100. What differs as a function of VOT is the likelihood of giving a 0 or 100 rating. In this case, at intermediate VOTs we see random fluctuations between the two endpoints, rather than responses clustered around an intermediate VAS value. Thus, this participant appears to have adopted a categorical approach. In contrast, Subject 8's responses closely follow the cue values of each stimulus, and variation is tightly clustered around the mean. Thus, subject 8's responses seem to reflect the gradient nature of the input.

To quantify individual differences, we fitted participants' VAS ratings using the rotated logistic function provided in Eq.1. Figure 8 shows the actual and fitted response functions for the two types of stimuli (labial and alveolar) across participants. Because the distribution of raw VAS slopes was positively skewed, we log-transformed values for analysis.

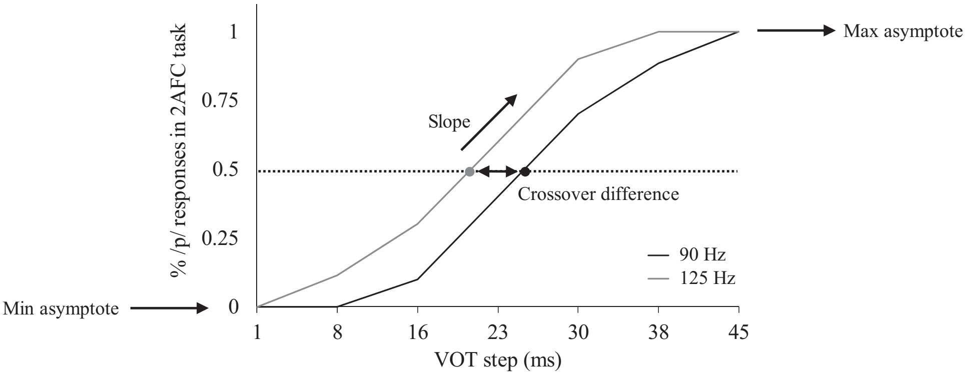


Figure 4. Hypothetical response curves in the 2AFC; (dark: low pitch; light: high pitch). 2AFC = 2-alternative forced choice; VOT = Voice Onset Time.

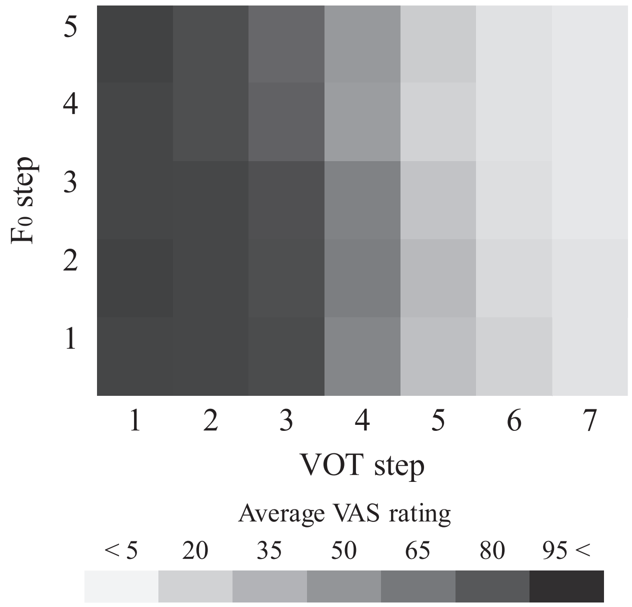


Figure 5. Visual analogue scaling responses by Voice Onset Time and F_0 steps.

We conducted an analysis of VAS scores as a function of stimulus type and place of articulation (PoA; see supplement S.3 for details). In brief, we found no significant effects of stimulus type or PoA on VAS slope. We also found evidence for higher use of F_0 for labial compared to alveolar stimuli.

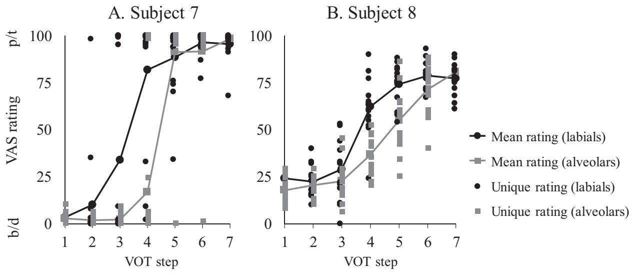


Figure 7. Sample VAS ratings per VOT and F_0 value exhibiting highly dissimilar patterns of noise; Subject 7 (left) responds categorically (close to the endpoints), but sometimes picks the wrong endpoint, whereas Subject 8 (right) closely maps his ratings to the VOT steps. VOT = Voice Onset Time; VAS = visual analogue scaling.

Descriptive Overview of 2AFC Data

The three participants that were excluded from the VAS analyses were also excluded from the 2AFC analyses. In addition, two additional participants were excluded due to technical issues, leaving 126 participants with data for this task.

Participants used both VOT and F_0 in the 2AFC task. They were more likely to categorize stimuli as /p/ (or /t/) when they had higher VOTs and higher F_0 values (see Figure 9). We fitted 2AFC data using Eq. 3. The distribution of 2AFC slopes was positively skewed, so these were log-transformed for analyses. Similarly, the distribution of raw crossover differences (our measure of F_0 use) was moderately positively skewed, so these were square-root transformed.

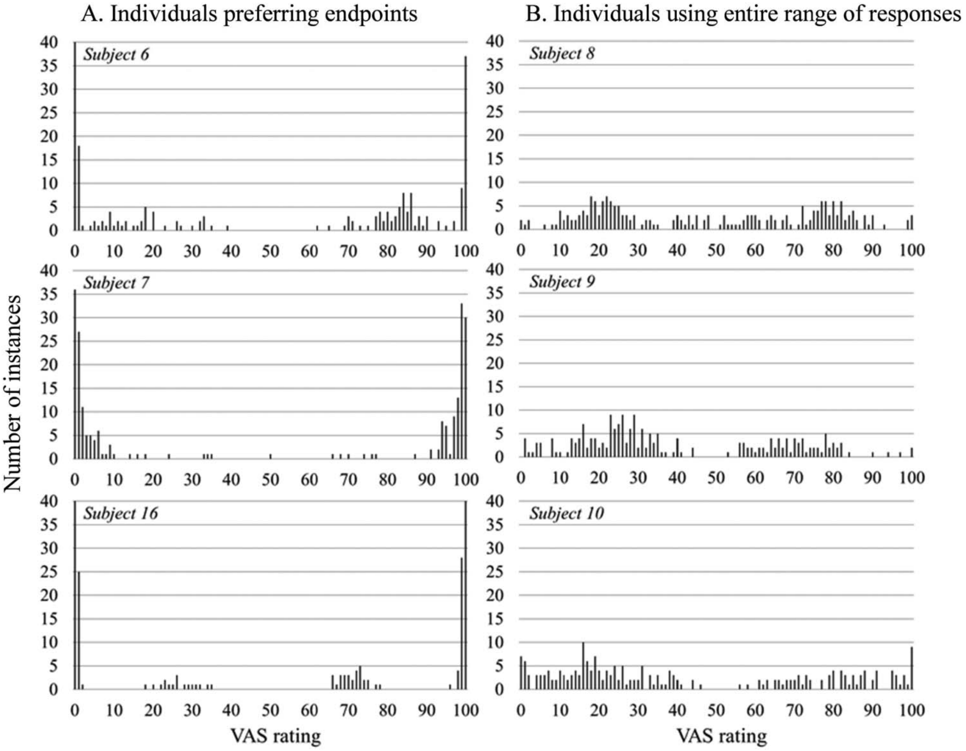


Figure 6. Histograms of sample individual visual analogue scaling responses.

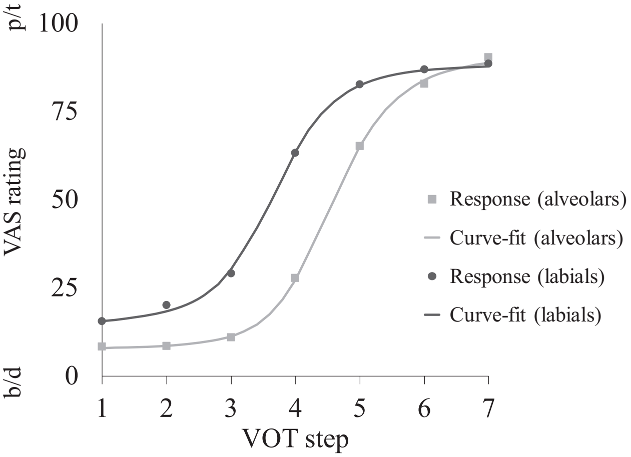


Figure 8. Actual and fitted VAS ratings (dark: labial; light: alveolar). VAS = visual analogue scaling; VOT = Voice Onset Time.

We analyzed 2AFC results by stimulus type and PoA (see supplement S.4 for details). In brief, we found no main effects of stimulus type or PoA on 2AFC slope. Second, similarly to the VAS task, listeners used F_0 more for labial stimuli and hardly at all for alveolars (Figure 9B).

Descriptive Analyses: Summary

Listeners were highly consistent across tasks in how they categorized stimuli (e.g., there was no main effect of stimulus type or PoA on slope in either task; and there was greater use of pitch information for labials in both tasks; see supplement S.5). Based on these results, we averaged slopes across PoA to compute a single slope estimate for each participant in each task. In addition, given the importance of multiple cue integration for our questions, only labial-initial stimuli were included in the analyses of F_0 use. More broadly, this close similarity in the pattern of effects between the VAS and 2AFC results validates the VAS task and is in line with a pattern of categorization that is relatively stable within individuals.

Individual Differences in Speech Perception

We next addressed our primary theoretical questions by examining how our speech perception measures were related to each other and to other measures.

Noise and gradiency in phoneme categorization. We first examined the relationship between VAS slope (categorization gradiency) and 2AFC slope (which may reflect categorization gradiency and/or internal noise in cue encoding). As slope was averaged across the two PoA, there were no repeated measurements, enabling us to use hierarchical regression to evaluate VAS slope as a predictor of 2AFC slope.

On the first level of the model (see Table 3), stimulus type was entered, contrast-coded into two variables, one comparing CVs to the other two (CV = 2; RW = −1; NW = −1), and the other comparing RWs to the other two (RW = 2; NW = −1; CV = −1). This explained 1.78% of the variance, $F(2, 117) = 1.06$, $p = .35$. On the second step, VAS slope was added to the model, which did not account for significantly additional variance ($R^2_{\text{change}} = .002$, $F_{\text{change}} < 1$). On the last step, we entered the VAS Slope × Stimulus Type interaction, which accounted for a marginally significant additional variance ($R^2_{\text{change}} = .048$, $F_{\text{change}}(5,114) = 2.96$, $p = .056$). To examine this interaction, we split the data by stimulus type; however, VAS slope did not account for a significant portion of the 2AFC slope variance in any of the subsets.[4]

This lack of correlation between 2AFC and VAS slope implies the two measures may reflect different aspects of speech categorization. As described, this may be because the 2AFC task is more sensitive to noise (in the encoding of cues), while the VAS reflects categorization gradiency. This is in line with Figure 7, which suggests that two subjects may have similar mean slopes in the VAS task despite large differences in the trial-to-trial noise around that mean. While the 2AFC task cannot assess this, the VAS task may be able to do so.

To test this hypothesis, we extracted a measure of noise in cue encoding from the VAS task using residuals. We first computed the difference between each VAS rating (on a trial-by-trial basis) and the predicted value from that subject's rotated logistic. We then computed the standard deviation of these residuals. This was done separately for each PoA and averaged to estimate the noise for each subject. The *SD* of the residuals in the VAS task was marginally correlated with 2AFC slope in the expected direction (negatively), $r = -.168$, $p = .063$. Listeners with shallower 2AFC slopes showed more noise in the VAS task. Interestingly, noise was weakly positively, though not significantly, correlated with VAS slope, $r = .120$, $p = .185$, suggesting that, if anything, listeners with higher gradiency (shallower VAS slope) are less noisy in their VAS ratings. This seems to agree with the sample results presented in Figure 7, as more gradient listeners tend to give ratings that more systematically reflect the stimulus characteristics.

Secondary cue use as a predictor of gradiency. Next, we examined whether gradiency in phoneme categorization was linked to multiple cue integration. As above, we used hierarchical regression with VAS slope as the dependent variable. Independent variables were stimulus type (coded as before) and F_0 use (the difference in 2AFC crossover points between low and high F_0). Only labial-initial stimuli were included. In the first level (see Table 4), stimulus type did not significantly account for variance in VAS slope, $R^2 = .014$, $F < 1$. In the second level, F_0 explained significant new variance, $\beta = -.296$; $R^2_{\text{change}} = .077$, $F_{\text{change}}(1, 116) = 9.87$, $p < .01$. On the last level, the F_0 use × Stimulus Type interaction did not significantly account for additional variance ($R^2_{\text{change}} = .024$, $F_{\text{change}}(2, 114) = 1.53$, $p = .220$). These results corroborate Kong and Edwards (2016): listeners with more phoneme categorization gradiency (shallower VAS slope)

[4] The lack of a significant relationship between the slopes for the two tasks raised the possibility that perhaps the VAS task is not related to more standard speech categorization measures. To confirm that the VAS task could provide good measures of basic aspects of speech perception (such as category boundary and secondary cue use), we also examined correlations between the crossover and F_0 use extracted from the two tasks (see supplement S.6). These show a robust relationship, supporting the validity of the VAS task.

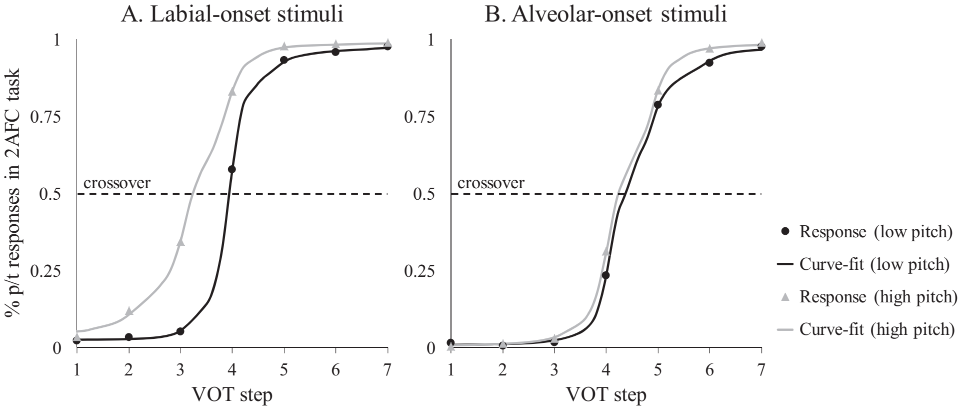


Figure 9. Actual and fitted 2AFC responses (dark: low pitch; light: high pitch). Figure depicts averages of fitted logistics, not fitted logistics of averages. 2AFC = 2-alternative forced choice; VOT = Voice Onset Time; VAS = visual analogue scaling.

showed greater use of F_0, suggesting a link between these aspects of speech perception.

Executive function and gradiency. Next we examined the relationship between executive function (EF) and categorization gradiency. Because the distribution of N-Back scores was positively skewed, while the distribution of the Trail Making task was moderately positively skewed, we used the log-transformed and square-rooted values respectively in all analyses.

We first estimated the correlations between EF measures. Flanker (inhibition) was not significantly correlated with either N-Back (working memory; $r = .01$) or Trail Making (executive function; $r = .12$). However, N-Back performance was weakly, but significantly, correlated with Trail Making ($r = .19$, $p < .05$).

We then conducted a series of regressions examining the relationship between phoneme categorization gradiency and EF (see Table 5). Three regressions were run, one for each EF measure-with VAS slope (averaged across PoA) as the dependent variable. In the first level of each model we entered stimulus type, and each EF measure was added in the second level.

As in prior analyses, stimulus type did not correlate with VAS slope. N-Back, however, explained a significant portion of the VAS slope variance, with higher N-Back scores significantly predicting shallower VAS slopes, $\beta = -.22$; $R^2_{\text{change}} = .045$, $F_{\text{change}}(1, 108) = 5.09$, $p < .05$ (Figure 10A). Trail Making did not predict VAS slope, $R^2_{\text{change}} = .014$, $F_{\text{change}}(2, 108) = 1.49$, $p = .23$ (Figure 10B), nor did Flanker, $R^2_{\text{change}} = .007$, $F_{\text{change}} < 1$, $p = .40$ (Figure 10C).

Executive function and multiple cue integration. The two prior analyses showed a relationship (a) between gradiency and cue integration and (b) between gradiency and N-Back performance (i.e., working memory). Given this, we next tested the possibility that the first correlation (gradiency and multiple cue integration) may be driven by a third factor, possibly EF. For example, greater working memory span may allow listeners to better maintain within-category information *and* better combine cues. We thus conducted hierarchical regressions with secondary cue use as the dependent variable. As above, three regression models were fitted-one for each EF measure with stimulus type at the first level, and an EF measure on the second.

Stimulus type had a significant effect on secondary cue use (see Table 6; see also supplement S.4), with significantly lower crossover differences (weaker use of F_0 as a secondary cue) for CV stimuli. On the second level, none of the EF measures was correlated with secondary cue use (*N*-Back: $R^2_{\text{change}} = .003$, $F_{\text{change}} < 1$, $p = .50$; Trail Making: $R^2_{\text{change}} = .006$, $F_{\text{change}} < 1$; Flanker: $R^2_{\text{change}} = .006$, $F_{\text{change}} < 1$). These results suggest that whatever the nature of the relationship between gradiency and multiple cue integration, it is unlikely to be driven by EF.

Speech-in-noise perception. Finally, we tested the hypothesis that maintaining within-category information may be beneficial for speech perception more generally. Speech recognition in noise

Table 3
Hierarchical Regression Steps: Predicting 2AFC Slope From VAS Slope

	Predictor	*B*	*SE*	β	R^2
Step 1	RW vs. others	.051	.036	.150	.018
	CV vs. others	.035	.036	.105	
Step 2	VAS slope	.088	.156	.052	.020
Step 3	VAS slope × RW vs others	−.215	.119	$-1.113^{\dagger}$	.068
	VAS slope × CV vs others	.092	.129	.484	

Note. RW = real words; CV = consonant-vowel; VAS = visual analogue scaling.
† $p < .1$.

Table 4
Hierarchical Regression Steps: Predicting VAS Slope From F_0 Use

	Predictor	*B*	*SE*	β	R^2
Step 1	RW vs others	.012	.027	.047	.014
	CV vs others	.034	.027	.133	
Step 2	F_0 use	−.341	.108	$-.296^{**}$	.091
Step 3	F_0 use × RW vs others	.077	.090	.098	.115
	F_0 use × CV vs others	−.092	.085	−.105	

Note. RW = real words; CV = consonant-vowel.
** $p < .01$.

Table 5
Hierarchical Regression Steps: Predicting VAS Slope From Executive Function Measures

Predictor		*B*	*SE*	β	R^2
Step 1	RW vs others	.011	.025	.051	.003
	CV vs others	−.001	.024	−.006	
Step 2a	N-Back	−.127	.057	−.215*	.048
Step 2b	Trail Making	−.062	.051	−.117	.017
Step 2c	Flanker	.052	.061	.082	.010

Note. RW = real words; CV = consonant-vowel.
* $p < .05$.

Table 6
Hierarchical Regression Steps: Predicting Secondary Cue Use From Executive Function Measures

Predictor		*B*	*SE*	β	R^2
Step 1	RW vs others	.004	.022	.017	.120
	CV vs others	−.071	.022	−.337**	
Step 2a	N-Back	.036	.053	.062	.123
Step 2b	Trail Making	−.043	.047	−.083	.126
Step 2c	Flanker	−.051	.055	−.084	.126

Note. RW = real words; CV = consonant-vowel.
** $p < .01$. *** $p < .001$.

(AzBio) was weakly negatively correlated with VAS slope ($r = -.16$), though this was only marginally significant ($p = .09$), suggesting more gradient VAS slopes might be beneficial for perceiving speech in noise. However, perception of speech in noise was significantly correlated with both N-Back performance ($r = .29$, $p < .01$) and Trail Making ($r = .29$, $p < .01$), and marginally correlated with Flanker ($r = .18$, $p = .055$). Thus, we assessed the relationship between gradiency and speech perception in noise after accounting for EF.

We again fitted hierarchical linear regressions with AzBio as the dependent variable. This time, in the first level, the three EF measures were entered simultaneously (see Step 1a in Table 7). These significantly explained 16.1% of the variance in AzBio performance, $F(3, 108) = 6.76$, $p < .001$,. Within this level, N-Back $\beta = .24$, $p < .01$, and Trail Making, $\beta = .22$, $p < .05$, were significant, while Flanker was marginal, $\beta = .15$, $p = .08$. As indicated by the beta coefficients (and Figure 11), higher scores in each EF measure predicted better AzBio performance.

In the second level, we added VAS slope, which did not account for significant new variance, $R^2_{\text{change}} = 0.01$, $\text{F}_{\text{change}}(1, 107) = 1.33$, $p = .24$. Finally, in the third level, we added two-way interactions between VAS slope and the three EF measures. None of the interaction terms accounted for significant new variance, $R^2_{\text{change}} = .007$, $\text{F}_{\text{change}} < 1$, $p = .84$. Thus, even though there was a marginally significant correlation between VAS slope and AzBio performance, when the three EF measures were included, this was no longer significant.

Next, we followed the reverse procedure, entering VAS slope in the first step (see Step 1b in Table 7). This was marginally significant, $\beta = -0.16$; $F(1, 110) = 2.85$, $p = .094$, explaining 2.5% of the variance. In the second level, we added the EF measures, which accounted for significant variance over and above VAS slope, $R^2_{\text{change}} = .14$, $\text{F}_{\text{change}}(3, 107) = 6.14$, $p < .001$. Thus, the relationship between gradiency and speech perception in noise may be largely due to individual differences in EF, with little unique variance attributable to gradiency.

Discussion

This study developed a novel way of assessing individual differences in speech categorization. The VAS task offers a unique approach to assessing gradiency of phoneme categorization, and contains a level of granularity that can robustly identify individual differences. While our most important finding was the correlation between phoneme categorization gradiency and cue integration, our correlational approach offers additional insights that are worth discussing before we turn to the implications of our primary finding.

Methodological Implications: The VAS and 2AFC Tasks

This study provides strong support for using VAS measures for assessing phoneme categorization. Monte Carlo simulations (supplement S.2) demonstrated that (a) the curve-fitting procedure was

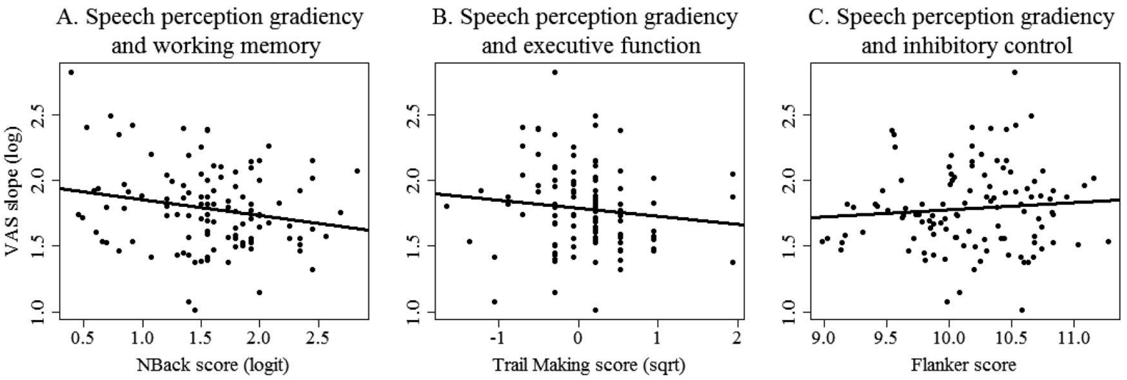


Figure 10. Scatter plots showing visual analogue scaling slope as a function of EF. (A) N-Back (working memory); (B) Trail Making (cognitive control); (C) Flanker (inhibition).

Table 7
Hierarchical Regression Steps: Predicting AzBio Score From Executive Function Measures and VAS Slope

	Predictor	*B*	*SE*	β	R^2
Step 1a	N-Back	.164	.062	.238*	.158
	Trail Making	.138	.056	.223*	
	Flanker	.113	.065	.155†	
Step 1b	VAS slope	−.184	.109	−.159	.025
Step 2	N-Back	.150	.063	.218*	.168
	Trail Making	.132	.056	.214*	
	Flanker	.120	.065	.165†	
	VAS slope	−.121	.105	−.105	
Step 3	VAS slope × N-Back	.140	.191	.073	.175
	VAS slope × Trail Making	−.025	.195	−.012	
	VAS slope × Flanker	−.109	.263	−.039	

Note. VAS = visual analogue scaling.
† $p < .1$. * $p < .05$.

unbiased and generated independent fits of gradiency and multiple cue integration, and (b) the fits accurately represented the underlying structure of the data even with as few as three repetitions per stimulus step.

In addition, when relating the pattern of effects obtained in the VAS and 2AFC task, we found the same stimulus-driven effects in both measures (supplement S.5), and that category boundaries and estimates of multiple cue use were correlated in the two tasks (supplement S.6). These findings (supplement S.3–6) suggested that our estimates of various effects are relatively stable across tasks. Therefore, these effects seem to reflect underlying aspects of the processing system that are somewhat stable for any given individual, validating our individual differences approach. Furthermore, the similarity between the VAS and the 2AFC results provides strong validation of the VAS task as an accurate and precise measure of phoneme categorization.

Given this, the lack of correlation between the VAS and 2AFC slopes was striking. We expected to find some correlation between the two, since both are thought to reflect at least partly the degree of gradiency in speech categorization. However, the 2AFC slope did not predict VAS slope. This could mean that these two measures assess different aspects of speech perception, perhaps more so than initially thought. That is, the 2AFC slope may largely reflect internal noise, rather than the gradiency of the response function (as does the VAS slope).

This study cannot speak to the exact locus of the noise that the 2AFC is tapping; it could be noise at a processing stage as early as the perception of acoustic cues, or it could be that cues are perceived accurately, but noise is introduced when they are maintained or as they are mapped to categories. In all these cases, the likely result would be greater inconsistency in participants' responses particularly near the boundary.

A number of arguments support the claim that the 2AFC task may reflect noise in how cues are encoded and used. For example, even if listeners make underlying probabilistic judgments about phonemes, when it comes to mapping this judgment to a response, the optimal strategy is to always choose the most likely response (rather than attempting to match the distribution of responses to the internal probability structure; Nearey & Hogan, 1986). Though it is unclear if some (or all) listeners do this, it suggests that the 2AFC slope may not necessarily reflect the underlying probabilistic mapping from cues to categories. In contrast to the 2AFC task, the VAS task may offer a unique window into this mapping, allowing us to extract information that is not accessible with other tasks. This is supported by our own analyses of trial-by-trial variation (i.e., noise) showing a markedly different relationship between noise and slopes from the 2AFC and VAS tasks. Interpreting these results cautiously, they suggest that variation in 2AFC slope may be more closely tied to *noise* in the system (higher noise = shallower slope), whereas VAS slope reflects the *gradiency* of speech categories.

This has a number of implications when we consider the use of phoneme categorization measures to assess populations with communication impairments. First, our findings seem to explain why gradient 2AFC responding is often associated with SLI and dyslexia, even as theoretical models and work with typical populations suggest a more gradient mode of responding is beneficial. In the former case, the 2AFC task is not tapping gradiency at all, but rather is tapping internal noise (which is likely increased in impaired listeners). As we show here, the VAS simultaneously taps both, with the slope of the average responding reflecting categorization gradiency and the *SD* of the residuals reflecting noise. A

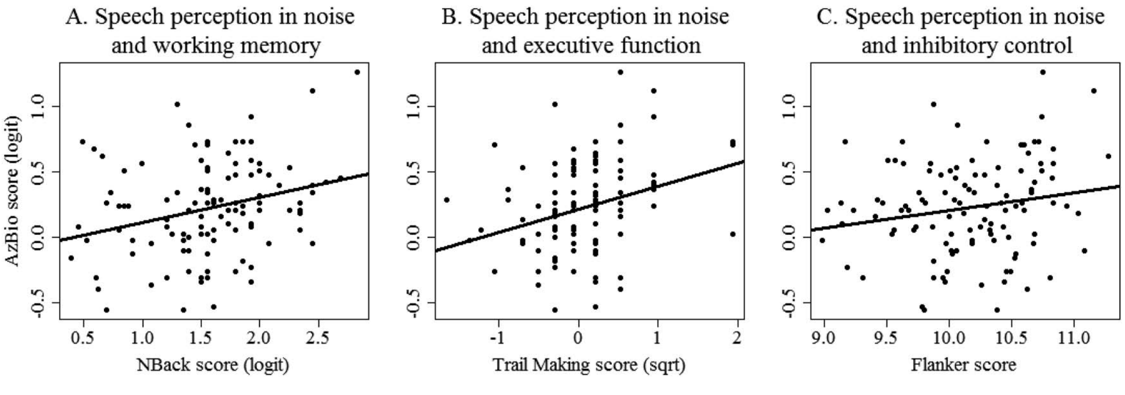


Figure 11. Scatter plots showing speech-in-noise perception (AzBio) as a function of EF. (A) N-Back (working memory). (B) Trail Making (cognitive control). (C) Flanker (inhibition). Note that an AzBio logit score of 0 (zero) corresponds to 50% accuracy.

combination of measures may thus offer more insight into the locus of perceptual impairments than traditional 2AFC measures, particularly when combined with online measures such as eye-tracking (cf., McMurray et al., 2014) that overcome other limitations of phoneme judgment tasks.

It is also helpful to consider how the 2AFC and VAS measures relate to other measures of gradiency. Many of the seminal studies supporting an underlying gradient form of speech categorization used a variant of the Visual World Paradigm (Clayards et al., 2008; McMurray et al., 2002, 2008). Here, gradiency is usually measured as the proportion of looks to a competitor (e.g., *bear* when the target is *pear*) as a function of step along a continuum such as VOT. Typically, as the step nears the boundary, competitor fixations increase linearly, suggesting sensitivity to within-category changes. This measure is computed relative to each participant's boundary, and only for trials on which participants click on the target. This allows us to extract a measure that may be less susceptible to the noise issues that appear with the 2AFC task. However, it is an open question whether this measure of gradiency may tap into the same underlying processes as those tapped by the VAS task.

Primary finding: Gradiency and cue integration. A critical result was that phoneme categorization gradiency was linked to multiple cue integration, such that greater gradiency predicted higher use of pitch-related information (see also Kong & Edwards, 2016). This finding is correlational, and therefore consistent with a number of possible causal accounts. First, multiple cue integration may allow for more gradient categorization. Under this view, the ability to integrate multiple cues may help listeners form a more precise graded estimate of speech categories. Alternatively, as we proposed, the causality may be reversed, with more gradiency helping listeners to be more sensitive to small differences in each cue, permitting better integration. Third, a gradient representation could help listeners avoid making a strong commitment on the basis of a single cue, allowing them to use both cues more effectively. Lastly, there could be a third factor that links the two. In this regard, we examined EF measures and found a relationship with gradiency for only the N-Back task, but no relationship between any EF measures and multiple cue integration. Additional factors of this sort—such as auditory acuity—should be considered in future research. Even though our study was not designed to distinguish between these mechanisms, it offers strong evidence for a link between these two aspects of speech perception, which remains to be clarified.

Links to other cognitive processes. Our findings show a potential link between working memory (*N*-Back) and participants' response pattern in the VAS task. One possibility for this correlation is that working memory mediates the relationship between gradiency and individuals' responses; there may be individuals who have gradient speech categories, but this gradient activation is not maintained all the way to the response stage due to working memory limitations. That is, the degree to which gradiency at the cue/phoneme level is reflected in an individual's response pattern may depend on their working memory span. Furthermore, measures that tap earlier stages of processing (e.g., ERPs, see Toscano et al., 2010), or earlier times in processing (e.g., eye-movements in the visual world paradigm: McMurray et al., 2002) may be less susceptible to working memory constraints, possibly explaining why these measures offer some of the strongest evidence for gradiency as a characterization of the modal listener.

Speech perception in noise. We predicted that higher gradiency would allow listeners to be more flexible in their interpretation of the signal and, thus, outperform listeners with lower levels of gradiency in a speech-in-noise task (AzBio sentences). Our results did not support this: gradiency was not a significant predictor of AzBio performance, which was, however, significantly predicted by our three EF measures (*N*-Back, Trail Making, and Flanker task).

The lack of correlation between gradiency and AzBio performance may reflect difficulties in linking laboratory measures of underlying speech perception processes (and cognitive processes more generally) to simple outcome measures. Such difficulty could arise from at least two sources. First, speech-in-noise perception may be more dependent on participants' level of motivation and effort than laboratory measures. This is supported by recent work on listening effort (Wu, Stangl, Zhang, Perkins, & Eilers, 2016; Zekveld & Kramer, 2014), which suggests that listeners put forth very low effort at low signal-to-noise ratios, they often appear to give up. Even though it is unlikely that in our study participants gave up in the AzBio task, the point being made here is that motivation may be a significant source of unwanted variability in these measures. Indeed, the significant correlations between our speech-in-noise measure and scores on the three executive function tasks may derive from a similar source. If so, this correlation may have little to do with speech perception.

Furthermore, while speech-in-noise perception is a standard assessment of speech perception accuracy, performance in such tasks may not be strongly affected by differences in categorization gradiency. As we describe in the introductory section of this article, theoretical arguments for gradiency are not typically framed in terms of speech-in-noise perception; rather, the motivation seems to derive from the demands of interpreting ambiguous acoustic cues, such as those related to anticipatory coarticulation, speaking rate, or speaker differences. Noise does not necessarily alter the cue values; rather, it masks the listeners' ability to detect them. Thus, this task may not properly target the functional problems that categorization gradiency is attempting to solve.

In a related vein, it may be the case that both gradient and categorical modes of responding are equally adaptive for solving the problem of speech perception in noise. That is, to the extent that differences in listeners' mode of categorization reflects a different weighting of different sorts of information (e.g., between acoustic or phonological representations in the Pisoni & Tash, 1974 model; or between dorsal and ventral stream processing in the Hickok & Poeppel, 2007 model), both sources of information may be equally useful for solving this problem (even as there are advantages of gradiency for other problems).

Gradiency and nongradiency in the categorization of speech sounds can both be advantageous in different ways. Therefore, in order to find the link that connects the underlying cognitive processes to a performance estimate, we need to use different measures of performance that are more closely tied to the theoretical view of speech perception that is being evaluated. Similar concerns may suggest the need to reconsider the way we evaluate speech perception tests used in a variety of different settings, including for clinical evaluations, so that they tap more into the underlying processes linked to our predictions.

Conclusions. We evaluated individual differences in phoneme categorization gradiency using a VAS task. This task, coupled with a novel set of statistical tools and substantial experimental extensions, allowed us for the first time to extract independent measures of speech categorization gradiency, multiple cue integration, and noise, at the individual level.

Our main results can be summarized as follows: First, we found substantial individual differences in how listeners categorize speech sounds, thus verifying the results by Kong and Edwards (2016) with a significantly larger sample. Second, we showed that differences in phoneme categorization gradiency seem to be theoretically independent from differences in the degree of internal noise in the encoding of acoustic cues and/or cue-to-phoneme mappings, and thus should not be confused with the traditional interpretation of shallow slopes as indicating noisier categorization of phonemes. Both categorization gradiency and such forms of noise contribute to speech perception, but may be tapped by different tasks. Third, differences in categorization gradiency are not epiphenomenal to other aspects of speech perception and appear to be linked to differences in multiple cue integration. The functional role of gradiency, however, is not yet clear, as the causal direction of this relationship remains to be defined. Fourth, we found only limited relationship between executive function and gradiency, suggesting that differences in categorization sharpness may derive from lower-level sources. Lastly, gradiency may be weakly related to speech perception in noise, but this seems to be modulated by executive function−related processes.

These results provide useful insights as to the mechanisms that subserve speech perception. Most importantly, they seem to stand in opposition to the commonly held assumption (see "Individual differences in phoneme categorization" section) that a sharp category boundary (and poor within-category discrimination) is the desired strategy for categorizing speech sounds efficiently and accurately. Overall, speech categorization is gradient, although to different degrees among listeners, and further work is necessary to reveal the sources of these differences and the consequences they have for spoken language comprehension.

References

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America, 106*(4 Pt 1), 2031–2039. http://dx.doi.org/10.1121/1.427949

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52,* 163–187. http://dx.doi.org/10.1016/0010-0277(94)90042-6

Blomert, L., & Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech. *Brain and Language, 89,* 21–26. http://dx.doi.org/10.1016/S0093-934X(03)00305-5

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience, 17,* 1353–1366. http://dx.doi.org/10.1162/0898929054985473

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 5.3) [Computer Program]. Retrieved from http://www.praat.org

Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., & Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology, 101,* 137–155. http://dx.doi.org/10.1016/j.jecp.2008.03.006

Carney, A. E., Widin, G., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961–970. http://dx.doi.org/10.1121/1.381590

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13,* 1428–1432. http://dx.doi.org/10.1038/nn.2641

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108,* 804–809. http://dx.doi.org/10.1016/j.cognition.2008.04.004

Coady, J. A., Evans, J. L., Mainela-Arnold, E., & Kluender, K. R. (2007). Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. *Journal of Speech, Language, and Hearing Research, 50,* 41–57. http://dx.doi.org/10.1044/1092-4388(2007/004)

Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research, 48,* 944–959. http://dx.doi.org/10.1044/1092-4388(2005/065)

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, *8,* 919–924. http://dx.doi.org/10.1097/00001756-199703030-00021

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 7126–7131. http://dx.doi.org/10.1073/pnas.1318738111

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149. http://dx.doi.org/10.3758/BF03203267

Frye, R. E., Fisher, J. M., Coty, A., Zarella, M., Liederman, J., & Halgren, E. (2007). Linear coding of voice onset time. *Journal of Cognitive Neuroscience, 19,* 1476–1487. http://dx.doi.org/10.1162/jocn.2007.19.9.1476

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

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology, 80*(11, Suppl. 3), S2–S6. http://dx.doi.org/10.1212/WNL.0b013e3182872e5f

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology*, *32,* 401–424. http://dx.doi.org/10.1016/0022-0965(81)90105-3

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. http://dx.doi.org/10.1037/0033-295X.105.2.251

Gow, D., Jr. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45,* 133–159. http://dx.doi.org/10.1006/jmla.2000.2764

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402. http://dx.doi.org/10.1038/nrn2113

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America, 109,* 748–763. http://dx.doi.org/10.1121/1.1337959

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099–3111. http://dx.doi.org/10.1121/1.411872

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77,* 30–60. http://dx.doi.org/10.1006/jecp.1999.2553

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358. http://dx.doi.org/10.1037/h0043688

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122,* 148–203. http://dx.doi.org/10.1037/a0038695

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

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59,* 40–57. http://dx.doi.org/10.1016/j.wocn.2016.08.006

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review, 23,* 1681–1712. http://dx.doi.org/10.3758/s13423-016-1049-y

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

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61,* 379–388. http://dx.doi.org/10.1037/h0049038

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4,* 187–196. http://dx.doi.org/10.1016/S1364-6613(00)01471-6

López-Zamora, M., Luque, J. L., Álvarez, C. J., & Cobos, P. L. (2012). Individual differences in categorical perception are related to sublexical/phonological processing in reading. *Scientific Studies of Reading, 16,* 443–456. http://dx.doi.org/10.1080/10888438.2011.588763

Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition, 142,* 345–350. http://dx.doi.org/10.1016/j.cognition.2015.05.009

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

McMurray, B., & Aslin, R. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6,* 203–229. http://dx.doi.org/10.1207/s15327078in0602_4

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1609–1631. http://dx.doi.org/10.1037/a0011747

McMurray, B., & Farris-Trimble, A. (2012). Emergent information-level coupling between perception and production. In A. C. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 369–395). Oxford, UK: Oxford University Press.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118,* 219–246. http://dx.doi.org/10.1037/a0022325

McMurray, B., & Jongman, A. (2016). What comes after /f/? Prediction in speech derives from data-explanatory processes. *Psychological Science, 27,* 43–52. http://dx.doi.org/10.1177/0956797615609578

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 57,* 1344–1362. http://dx.doi.org/10.1044/2014_JSLHR-L-13-0196

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60,* 1–39. http://dx.doi.org/10.1016/j.cogpsych.2009.06.003

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42. http://dx.doi.org/10.1016/S0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language, 60,* 65–91. http://dx.doi.org/10.1016/j.jml.2008.07.002

Messaoud-Galusi, S., Hazan, V., & Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals? *Journal of Speech, Language, and Hearing Research, 54,* 1682–1701. http://dx.doi.org/10.1044/1092-4388(2011/09-0261)

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43*(1–3), 106–115. http://dx.doi.org/10.1159/000261764

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and cognitive processes, 12*(5-6), 865–870. http://dx.doi.org/10.1080/016909697386754

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46,* 505–512. http://dx.doi.org/10.3758/BF03208147

Moberly, A. C., Lowenstein, J. H., & Nittrouer, S. (2016). Word recognition variability with cochlear implants: "Perceptual attention" versus "auditory sensitivity". *Ear and Hearing, 37,* 14–26. http://dx.doi.org/10.1097/AUD.0000000000000204

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science, 20,* 895–903. http://dx.doi.org/10.1111/j.1467-9280.2009.02380.x

Nearey, T., & Hogan, J. (1986). Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves. In J. J. Ohala & J. J. J (Eds.), *Experimental Phonology* (pp. 141–146). Orlando, FL: Academic Press.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172–191. http://dx.doi.org/10.1037/0033-295X.85.3.172

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America, 99,* 1718–1725. http://dx.doi.org/10.1121/1.414696

Ojemann, G., Ojemann, J., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery, 71,* 316–326. http://dx.doi.org/10.3171/jns.1989.71.3.0316

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., . . . Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12,* 1038–1055. http://dx.doi.org/10.1162/08989290051137567

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55,* 328–333. http://dx.doi.org/10.1121/1.1914506

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285–290. http://dx.doi.org/10.3758/BF03213946

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81–110. http://dx.doi.org/10.1037/0033-2909.92.1.81

Repp, B. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practice, 10,* 243–335. http://dx.doi.org/10.1016/B978-0-12-608610-2.50012-1

Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science, 12,* 753–767. http://dx.doi.org/10.1111/j.1467-7687.2009.00806.x

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71,* 145–163. http://dx.doi.org/10.1016/j.jml.2013.11.002

Sams, M., Aulanko, R., Aaltonen, O., & Näätänen, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience, 2,* 344–357. http://dx.doi.org/10.1162/jocn.1990.2.4.344

Schellinger, S. K., Edwards, J., Munson, B., & Beckman, M. E. (2008, November). *Assessment of children's speech production 1: Transcription categories and listener expectations*. Poster presented at the 2008 ASHA Convention, Chicago, IL.

Schellinger, S. K., Munson, B., & Edwards, J. (2017). Gradient perception of children's productions of /s/ and /θ/: A comparative study of rating methods. *Clinical Linguistics & Phonetics, 31,* 80–103. http://dx.doi.org/10.1080/02699206.2016.1205665

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

Schouten, M. E. H., & van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America, 92,* 1841–1855. http://dx.doi.org/10.1121/1.403841

Serniclaes, W. (2006). Allophonic perception in developmental dyslexia: Origin, reliability and implications of the categorical perception deficit. *Written Language and Literacy, 9,* 135–152. http://dx.doi.org/10.1075/wll.9.1.09ser

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research, 44,* 384–399. http://dx.doi.org/10.1044/1092-4388(2001/032)

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology, 87,* 336–361. http://dx.doi.org/10.1016/j.jecp.2004.02.001

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98*(2), B35–B44. http://dx.doi.org/10.1016/j.cognition.2005.03.002

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., . . . Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and Hearing, 33,* 112–117. http://dx.doi.org/10.1097/AUD.0b013e31822c2549

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62,* 435–448. http://dx.doi.org/10.1121/1.381544

Sussman, J. E. (1993). Perception of formant transition cues to place of articulation in children with language impairments. *Journal of Speech Language and Hearing Research, 36,* 1286–1299. http://dx.doi.org/10.1044/jshr.3606.1286

Tombaugh, T. N. (2004). Trail making test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19,* 203–214. http://dx.doi.org/10.1016/S0887-6177(03)00039-8

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21,* 1532–1540. http://dx.doi.org/10.1177/0956797610384142

Urberg-Carlson, K., Kaiser, E., & Munson, B. (2008, November). *Assessment of children's speech production 2: Testing gradient measures of children's productions*. Poster presented at the 2008 ASHA Convention, Chicago, IL.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62,* 1297–1311. http://dx.doi.org/10.3758/BF03212131

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41,* 48–61. http://dx.doi.org/10.1037/h0084150

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research, 56,* 1097–1107. http://dx.doi.org/10.1044/1092-4388(2012/12-0086)

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *Journal of the Acoustical Society of America, 137,* 1430–1442. http://dx.doi.org/10.1121/1.4908308

Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric Functions of Dual-Task Paradigms for Measuring Listening Effort. *Ear and Hearing, 37,* 660–670. http://dx.doi.org/10.1097/AUD.0000000000000335

Yeni-Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America, 70,* 966–975. http://dx.doi.org/10.1121/1.387031

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology, 51,* 277–284. http://dx.doi.org/10.1111/psyp.12151

Received March 7, 2016
Revision received January 29, 2017
Accepted January 30, 2017 ■