Published in final edited form as:
Psychol Sci. 2010 October ; 21(10): 1532–1540. doi:10.1177/0956797610384142.

Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech

Joseph C. Toscano,
Dept. Of Psychology University of Iowa

Bob McMurray,
Dept. Of Psychology University of Iowa

Joel Dennhardt, and
University of Tennessee at Memphis

Steven. J. Luck
Dept. of Psychology and Center for Mind & Brain University of California, Davis

Abstract

Speech sounds are highly variable, yet listeners readily extract information from them and transform continuous acoustic signals into meaningful categories during language comprehension. A central question is whether perceptual encoding captures continuous acoustic detail in a one-to-one fashion or whether it is affected by categories. We addressed this in an event-related potential (ERP) experiment in which listeners categorized spoken words that varied along a continuous acoustic dimension (voice onset time; VOT) in an auditory oddball task. We found that VOT effects were present through a late stage of perceptual processing (N1 component, ca. 100 ms poststimulus) and were independent of categories. In addition, effects of within-category differences in VOT were present at a post-perceptual categorization stage (P3 component, ca. 450 ms poststimulus). Thus, at perceptual levels, acoustic information is encoded continuously, independent of phonological information. Further, at phonological levels, fine-grained acoustic differences are preserved along with category information.

Keywords

Speech Perception; Language; Electrophysiology; Cognitive Neuroscience; Cognitive Processes

The acoustics of speech are characterized by immense variability. Individual speakers differ in how they produce words, and even the same speaker will produce different acoustic patterns across repetitions of a word. Despite this variability, listeners can accurately recognize speech. Thus, a central question in spoken language comprehension is how listeners transform variable acoustic signals into less variable, linguistically-meaningful categories. This process is fundamental for basic language processing, but is also relevant to other areas, such as language and reading impairment (Thibodeau & Sussman, 1979; Werker & Tees, 1987) and automatic speech recognition.

Corresponding Author: Joseph Toscano Dept. of Psychology E11 SSH University of Iowa Iowa City, IA 52242 joseph-toscano@uiowa.edu 319-335-0692 (voice) 319-335-0191 (fax).

Speech perception has been framed in terms of two levels of processing (Pisoni, 1973): the perceptual encoding[1] of continuous acoustic cues and the subsequent mapping of this information onto categories like phonemes or words. Theories of speech perception differ in the nature of representations at both levels and the transformations that mediate them (Oden & Massaro, 1978; Liberman & Mattingly, 1985; McClelland & Elman, 1986; Goldinger, 1998). Historically, a dominant question was whether perception is graded or categorical (discrete or nonlinear) with respect to the continuous input (Liberman, Harris, Hoffman, & Griffith, 1957; Schouten, Gerrits, & Van Hessen, 2003). Such discreteness could arise from several sources: an inherent, nonlinear encoding of speech into articulatory gestures (Liberman & Mattingly, 1985); the learned influence of phonological categories (Anderson, Silverstein, Ritz, & Jones, 1977); or discontinuities in low-level auditory processing (Sinex, MacDonald, & Mott, 1991; Kuhl & Miller, 1978).

If perception is nonlinear in one of these ways, listeners will be less sensitive (or completely insensitive) to differences within a category than to differences between categories. Consider voice onset time (VOT), the time difference between the release of constriction and the onset of voicing. VOT leaves an acoustic trace that serves as a continuous cue distinguishing voiced (/b,d,g/) from voiceless (/p,t,k/) stops. If perception of VOT is categorical, then VOTs between 0 and 20 ms (/b/) may be encoded as more similar to each other than to VOTs greater than 20 ms (/p/), even if the acoustic distance between them is the same.

Early behavioral work suggested that perception is categorical: listeners are poor at discriminating acoustic differences within the same category and good at equivalent distances spanning a boundary (Liberman et al., 1957; Repp, 1984). This supported a view that early perceptual processes encode speech in terms of categories and abstract away from fine-grained detail in the signal. Subsequent research challenged this latter claim, demonstrating that within-category differences are discriminable (Pisoni & Tash, 1974; Carney, Widin & Viemeister, 1977; Massaro & Cohen, 1983) and that such differences are meaningful: phoneme (Miller, 1997; McMurray, Aslin, Tanenhaus, Spivey & Subik, 2008) and lexical categories (Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus & Aslin, 2002) show a graded structure that is sensitive to within-category distinctions.

Thus, the issue of within-category sensitivity has been settled: listeners are sensitive to fine-grained acoustic information. However, we do not know whether *perceptual encoding* itself is linear or nonlinear, a critical distinction for determining what information listeners have access to when dealing with variability in the speech signal. To assess this, we must examine the perceptual representations of acoustic cues. These may vary linearly with the acoustic input, and category boundaries could be established at a later stage. Alternatively, perceptual encoding may be nonlinear for one of the reasons discussed above.

This question is difficult to answer with behavioral techniques since they reflect the combined influence of perceptual and categorization processes and are sensitive to task characteristics (Carney et al., 1977; Massaro & Cohen, 1983; Gerrits & Schouten, 2004). Some studies have shown that discrimination is independent of categories, but this has only been observed with unnatural tasks, not in situations that reflect real-world language processing (Massaro & Cohen, 1983; Schouten et al., 2003; Gerrits & Schouten, 2004).

[1]We use the term perceptual encoding to refer to the process by which continuous acoustic information (e.g., a particular VOT value) is converted to a representation usable by the system. Categorization refers to the process that uses the information provided by perceptual encoding to identify a phonological category.

Measurements of neural activity may allow us to observe perceptual representations more directly. Blumstein, Myers, and Rissman (2003), for example, used fMRI to assess within-category sensitivity, but differences were examined with respect to the phonological category rather than raw VOT. Thus, this does not address the question of perceptual encoding.

The event-related potential (ERP) technique offers a useful tool for isolating perceptual activity from categorization during real-time processing. Numerous ERP experiments have used the mismatch negativity (MMN) as a measure of change detection or discrimination (Näätänen & Picton, 1987; Pratt, in press), but they offer conflicting results. Several studies suggest that phonological categories influence the MMN (Sams et al., 1990; Dehaene-Lambertz, 1997; Sharma & Dorman, 1999; Phillips et al., 2000 [using the equivalent MEG component]), while other studies have suggested that it does not (Sharma et al., 1993; Joanisse et al., 2007). More importantly, the MMN is defined as a difference in responses between a rare stimulus and a frequent stimulus, requiring the use of contrived tasks that make it difficult to assess how each stimulus is independently represented.

Neurophysiological work has also examined responses to individual stimuli, allowing a better comparison to natural language processes. Many of these studies suggest discontinuous encoding of continuous cues (Steinschneider, Volkov, Noh, Garell, & Howard, 1999; Sharma & Dorman, 1999; Sharma, Marsh, & Dorman, 2000). Studies measuring the auditory N1 ERP component have found a single peak for short VOTs and a double peak for long VOTs. However, if the first peak is driven by the release burst and the second peak by voicing onset, the two peaks may merge when they occur close in time (i.e., at short VOTs; Sharma et al., 2000). Frye et al. (2007) report a single peak for the M100 MEG component (an analogue of the N1) for both short and long VOTs, consistent with the prediction that encoding is continuous. However, they only examined four stimulus conditions, making it difficult to assess whether the response is linear across the entire VOT continuum and to rule out variation between participants' categories as a source of this result. The N1 may be sensitive to within-category differences, but observing this may be difficult in stimuli with a high-amplitude burst.

Thus, we do not know whether there is a level of processing at which speech cues are encoded linearly. Data showing sensitivity to fine-grained detail at later stages do not address this issue per se, and the neural evidence of early processing has been inconclusive.

We assessed sensitivity at perceptual levels to determine if a linear relationship between an acoustic cue (VOT) and brain responses could be found. We used the fronto-central auditory N1, which has been shown to respond to a wide range of stimulus types (Näätänen & Picton, 1987), as a measure of perceptual-level processing. This component is generated in auditory cortex within Heschl's gyrus approximately 50 ms after the initial primary auditory cortex response (Pratt, in press) and, thus, originates late in perceptual processing but early in language processing. Importantly, our stimuli do not contain large bursts, minimizing the problem of overlapping N1s described by Sharma et al. (2000).

We also assessed gradiency at the level of phonological categories using the P3 component, which has been shown to reflect categorization in a number of domains, including speech (Maiste, Wiens, Hunt, Scherg, & Picton, 1995), and should reflect phonological categorization of the stimuli. This was intended to confirm the predictions from behavioral experiments suggesting that fine-grained detail is preserved later in language processing.

If listeners are sensitive to within-category acoustic variation, we should see this sensitivity in both the N1 and P3 components. More importantly, if listeners encode perceptual

information linearly at early stages of processing, we predict that the N1 will not show effects of phonological category information or auditory discontinuities.

Methods

Design

Participants heard four equi-probable words (*beach, peach, dart, tart*) over Sennheiser 570 headphones while we recorded ERPs from scalp electrodes. VOT was manipulated between 0 ms (prototypical for *beach* and *dart*) and 40 ms (prototypical for *peach* and *tart*) in nine steps. Each word was designated the target in different blocks. Participants were instructed to press one button for the target word and a different button for any other stimulus. Thus, approximately 25% of the stimuli were categorized as targets (sufficient to produce a P3 wave), but the actual probability of the target category depended on the participant's VOT boundary and generally varied between 17% and 33% across participants and continua (Fig. S1).

Behavioral Task

In each of the four blocks, one of the two continua was task-relevant and the other was task-irrelevant, depending on which word was designated the target for that block (e.g., when *dart* was the target, *dart-tart* was task-relevant and *beach-peach* was task-irrelevant). Block order varied between participants with the requirement that the same continuum could not be task-relevant on successive blocks.

A gamepad recorded behavioral responses. Participants pressed one button with either their left or right hand (alternated between participants) to make a target response and another button with the opposite hand to make a non-target response. Eighteen practice trials were presented at the beginning of each block. Participants were given the opportunity to take a short break every 35 trials, and there was a longer break halfway through the experiment. A total of 630 trials were presented (not including practice trials), and each of the nine steps of the two continua was presented approximately 35 times per block.[2]

After the main experiment, participants performed a 2AFC labeling task in which they categorized each token of the continuum as starting with "B" or "P" for the *beach/peach* continuum and "D" or "T" for the *dart/tart* continuum. Each continuum was presented in a separate block, and stimuli were randomly presented six times within each block. We computed participants' boundaries by fitting logistic functions to these data.

Participants

Participants were recruited from the University of Iowa community, provided informed consent, and were compensated $8 per hour. The final sample included 17 participants (12 female; approximate age range: 18-30 years). Data from three of these were excluded from the P3 analyses due to problems with the behavioral task that was run at the end of the experiment. All participants were included in the N1 analyses, since they did not rely on these boundaries.

Stimuli

Stimuli were constructed using the KlattWorks front-end (McMurray, 2009) to the Klatt (1980) synthesizer. Stimuli began with a 5 ms burst of low-amplitude frication. To create the

[2]A bug in the randomization software prevented this from being perfectly equivalent across conditions. The average standard deviation in the number of repetitions was 4.3.

VOT continua, AV (amplitude of voicing) was cut back in 5 ms increments and replaced with 60 dB of AH (aspiration). All other parameters were constant across VOT steps. For the *beach-peach* continuum, F1, F2, and F3 transitions had rising frequencies, and for the *dart-tart* continuum, F2 and F3 transitions had falling frequencies and F1 had a rising frequency. Formant frequencies for vowels were based on spectrographic analysis of natural tokens.

EEG recordings

ERPs were recorded from standard electrode sites over both hemispheres (International 10/20 System sites F_3, F_4, F_Z, C_3, C_z, C_4, P_3, P_z, P_4, T_3, T_4, T_5, and T_6), referenced to the left mastoid during recording and re-referenced offline to the average of the left and right mastoids. Horizontal electrooculogram (EOG) recordings were obtained using electrodes located 1 cm lateral to the external canthi for each eye, and the vertical EOG was recorded using an electrode beneath the left eye. Impedance was 5 kΩ or less at all sites. The signal was amplified using a Grass Model 15 Neurodata Amplifier System with a notch filter at 60 Hz, a high-pass filter at 0.01 Hz, and a low-pass filter at 100 Hz. Data were digitized at 250 Hz.

Data Processing

Data were processed using the EEGLAB toolbox for MATLAB (Delorme & Makeig, 2004). Trials containing ocular artifacts, movement artifacts, or amplifier saturation were rejected. Artifact rejection was performed in two stages. In the first stage, trials were automatically marked if they contained voltages that exceeded a threshold of 75 μV in any of the EOG channels or 150 μV in any of the EEG channels. The data were then visually inspected to reject trials with any additional artifacts. Baselines for each epoch were computed as the mean voltage 200 ms before the onset of the stimulus.

Behavioral responses

Participants' behavioral responses showed standard categorization functions for both continua, though boundaries were affected by which target the participant was monitoring for (Fig. 1A). Participants' responses in the labeling task performed after ERP recording also showed standard categorization functions (Fig. 1B). (See Results in the supporting information online.) N1 amplitude

N1 amplitudes were measured as the mean voltage from 75-125 ms poststimulus from the average of the three frontal channels (Figs. 2A and 2B). N1 amplitude decreased with increasing VOT, and this effect was observed for both the relevant and irrelevant continua.

The data were analyzed using a linear mixed-effects model (LMM) using the lme4 package in R with the within-subject factors of VOT, stimulus continuum [*beach/peach* or *dart/tart*], target voicing [*voiced or voiceless*], and task relevance [*relevant* or *irrelevant*] as fixed effects (see Results in the supporting materials online for ANOVAs). In this and subsequent LMM analyses, participant was entered as a random effect, and the Markov Chain Monte Carlo (MCMC) procedure was used to estimate p-values for the coefficients. The main effect of VOT was significant ($b=0.215$ $p_{MCMC}<0.001$)[3], confirming the prediction that VOT affects the magnitude of the N1. The main effect of stimulus continuum was also significant ($b=0.744$, $p_{MCMC}<0.001$), with larger N1 amplitudes for *beach/peach* than *dart/tart*. Thus, the N1 encodes not only VOT, but also differences in acoustic information more broadly. All other main effects and interactions were non-significant.

[3]The model coefficients reported here are unstandardized, so the numbers reflect values in μV per unit of the factor.

We next asked whether the effect of VOT on N1 amplitude could be fit just as well by a categorical model in which the category boundary varies across individuals (which could produce results mimicking linearity across participants). We directly compared two mixed-effects models relating N1 amplitude to VOT (similar to McMurray, Tanenhaus, & Aslin, 2009): a linear model defined by two parameters (slope and intercept); and a categorical model, defined as a step function with three parameters (the lower bound, the upper bound, and the crossover point). In both models, parameters were fit to each participant's data to ensure that linear results were not an artifact of averaging.

The linear model provided a better overall fit, as measured both by mean R^2 values (linear: 0.430; categorical: 0.343) and the Bayesian Information Criterion (BIC) (linear: 645.2; categorical: 745.6). BIC scores for individual participants favored the linear model for 16 of 17 participants (binomial test: $p<0.001$). Thus, even though the categorical model had more free parameters, the linear model provided a better fit, suggesting that responses to acoustic cues are predominantly linear.[4]

The final analysis was intended to examine potential influences of phonological categories on the N1. That is, for a given VOT, does the N1 differ on the basis of how it was classified? To do this, we restricted the dataset to include only trials in which participants made a *target* response. On some of these trials the participant was supposed to make a target response when they heard a voiceless token, and on others when they heard a voiced token. Thus, if the participant's category for the stimulus influenced the relationship between N1 and VOT, we might observe a main effect of target voicing or its interaction with VOT. This also allows us to confirm that variation in category membership could not account for the pattern observed above, since all stimuli were identified the same for a given target voicing condition. The use of only trials with *target* responses meant that some conditions would include more trials than others (e.g. there were few trials with VOTs of 40 ms when participants indicated a voiced target), so each data point was weighted by the number of trials in that condition.[5] Fig. 2C shows N1 amplitude for the different conditions in this dataset.

An LMM analysis with VOT and target voicing as fixed factors showed a significant main effect of VOT ($b=0.317$, $p_{MCMC}<0.001$); neither target voicing nor the interaction were significant. Thus, there was no evidence to support the influence of phonological category information on N1 amplitude.

P3 amplitude

P3 amplitudes were measured from the average of the three parietal channels by computing the mean voltage between 300 and 800 ms after stimulus onset. Fig. 3 shows the ERP waveform as a function of distance from the target along the task-relevant continuum (e.g., from *beach* to *peach* when *beach* was the target; Panel A) and along the task-irrelevant continuum (e.g., from *beach* to *peach* when *dart* was the target; Panel B). P3 amplitude was larger for the target end of the relevant continuum than for the non-target end, regardless of which word served as the target, and did not vary along the irrelevant continuum.

As with the N1, P3 amplitude was analyzed using an LMM with VOT, stimulus continuum, target voicing, and task relevance as fixed factors. There was a significant main effect of task relevance ($b=2.92$, $p_{MCMC}<0.001$), which was due to the presence of a P3 for the task-relevant but not the irrelevant continuum. There was a significant target voicing x VOT interaction ($b=0.373$, $p_{MCMC}<0.001$), with a larger P3 for short VOTs for *voiced* targets and

[4]The linear model also showed a better fit for each stimulus continuum individually.
[5]The unweighted model produced the same pattern of results.

a larger P3 for long VOTs for *voiceless* targets. This is consistent with the prediction that the P3 is sensitive to the category of the target. The stimulus continuum x task relevance interaction was significant (b=−0.893, $p_{MCMC}<0.001$), with a larger effect of task relevance for the *beach/peach* than the *dart/tart* continuum. A follow-up analysis found a significant effect of task relevance for both continua (*beach/peach*: b=3.36, $p_{MCMC}<0.001$; *dart/tart*: b=2.47, $p_{MCMC}<0.05$).

The target voicing x task relevance x VOT interaction in the original analysis was also significant (b=0.599, $p_{MCMC}<0.001$), since the interaction between VOT and target voicing was observed only for the relevant continuum. A follow-up analysis including only the task relevant trials showed a significant VOT x target voicing interaction (b=0.672, $p_{MCMC}<0.001$), confirming the predicted effect; other effects were non-significant. All other main effects and interactions in the original analysis were non-significant.

We next asked whether the P3 exhibited a gradient pattern within a category, controlling for the possibility that this was an artifact of averaging across different category boundaries. We computed VOT values relative to each participant's category boundary, or rVOT, and excluded all trials that a given participant categorized as being non-target (McMurray et al., 2008). Thus, the participant's behavioral response indicated that all of these VOT steps fell within the same category. These analyses considered only the relevant continuum, since no P3 was observed for the task-irrelevant one. Fig. 3C shows the ERP waveform for the rVOT closest to the boundary and three within-category rVOT steps away from the boundary and toward the target endpoint of the continuum.[6] Fig. 3D shows mean P3 amplitudes for these data.

Because LMMs assume linear effects, we excluded the most extreme rVOT values, as we did not expect much variation in P3 amplitude for these stimuli given that they are located well within participants' categories. Thus, we only analyzed trials in which the absolute value of rVOT was less than 4.5, excluding 18 out of 252 data points.[7] An LMM analysis with absolute valued rVOT, stimulus continuum, and target voicing as fixed effects showed a significant effect of rVOT (b=0.508, $p_{MCMC}<0.001$) with smaller P3 amplitude as rVOT approached the category boundary; other main effects and interactions were non-significant. Thus, listeners showed gradient sensitivity to VOT relative to their own boundary within each phonological category.

Discussion

These results indicate that (1) both N1 and P3 amplitude reflect listeners' sensitivity to fine-grained differences in VOT, and (2) while P3 amplitude is influenced by phonological categories, N1 amplitude is not. The N1 shows a one-to-one correspondence with VOT even when participants indicate that stimuli belong to different phonological categories and when differences in individual category boundaries are accounted for. Further, this effect is not specific to VOT, as N1 amplitude varies with place of articulation as well.

This constitutes strong evidence that non-phonological representations of speech are maintained until late (>100 ms) stages of perceptual processing and that listeners encode acoustic cues linearly prior to categorization. This fits with the hypothesis that speech perception is fundamentally continuous (Massaro & Cohen, 1983) and that effects of

[6]Some participants had only three steps on one side of the category boundary for one continuum. Thus, each data point contains data from every participant, but some participants contributed more data to the ±4 conditions than others.

[7]An analysis including the extreme rVOT values still produced a marginal main effect of rVOT (b=0.249 , $p_{MCMC}\approx0.054$) with no other significant effects.

phonological information are a product of categorization and task demands, not perceptual encoding.

This result contrasts with earlier claims that the morphology of the N1 reflects categorical perception (Sharma & Dorman, 1999; Sharma et al., 2000). However, as noted above, differences in the construction of stimuli allowed us to observe effects that may have been masked in previous studies. It also contrasts with work suggesting an auditory discontinuity in VOT encoding near the phonological boundary (Kuhl & Miller, 1978; Sinex et al., 1991), which would lead to a non-linearity in perceptual encoding. However, the evidence for such a fixed discontinuity is mixed, with estimates of its location ranging from 20 to 70 ms, depending on the specific characteristics of the stimuli and range of VOTs tested (Ohlemiller et al., 1999). Further, given that listeners must use VOT information flexibly in both developmental time (since the VOT categories for a particular language must be learned) and real-time during speech processing (e.g., because of variation in speaking rate), such an auditory discontinuity could make speech perception a much more difficult task.

The P3 results demonstrated that graded acoustic detail is also preserved at post-perceptual levels. The P3 occurs too late (ca. 450 ms) to be an indicator of phonological processing per se (though we refer to phonological categories here, since they were the relevant distinction in this task). Thus, acoustic detail is maintained even at post-phonological stages, consistent with behavioral and neuroimaging work showing graded phonetic categorization (Andruski et al., 1994; McMurray et al., 2002, 2009; Blumstein et al., 2003).

These results offer a basis for examining the nature of early acoustic cue processing and current work is extending this approach to other cues. The results may also have practical implications. Work on specific language impairment and dyslexia has suggested that impaired listeners show less categorical perception (Thibodeau & Sussman, 1979; Werker & Tees, 1987). However, if perceptual encoding is continuous and categorical effects emerge as an effect of task differences, we may need to use other measures as a benchmark for understanding speech perception in these groups (e.g., McMurray, Samelson, Lee, & Tomblin, 2010).

Together, the N1 and P3 results support a model of spoken word recognition in which perceptual processing is continuous and categorization is graded. More importantly, this linear encoding of the acoustic input is exactly what is needed for processes that use such detail to facilitate language comprehension (McMurray et al., 2009; Goldinger, 1998). This supports an emerging view that language processing is based on continuous and probabilistic information at multiple levels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Margot Schneider and Kelli Kean for assistance with data processing and Susan Wagner Cook for help with the statistical analyses. This research was supported by NIH DC008089 to B.M.

References

Anderson JA, Silverstein JW, Ritz SA, Jones RS. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. Psychological Review. 1977; 84:413–451.

Andruski JE, Blumstein SE, Burton M. The effect of subphonetic differences on lexical access. Cognition. 1994; 52:163–187. [PubMed: 7956004]

Blumstein SE, Myers EB, Rissman J. The perception of voice onset time: an fMRI investigation of phonetic category structure. Journal of Cognitive Neuroscience. 2003; 17:1353–1366. [PubMed: 16197689]

Carney AE, Widin GP, Viemeister NF. Non categorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America. 1977; 62:961–970. [PubMed: 908791]

Dehaene-Lambertz G. Electrophysiological correlates of categorical phoneme perception in adults. NeuroReport. 1997; 8:919–924. [PubMed: 9141065]

Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004; 134:9–21. [PubMed: 15102499]

Frye RE, Fisher J. McGraw, Coty A, Zarella M, Liederman J, Halgren E. Linear coding of voice onset time. Journal of Cognitive Neuroscience. 2007; 19:1476–1487. [PubMed: 17714009]

Gerrits E, Schouten MEH. Categorical perception depends on the discrimination task. Perception & Psychophysics. 2004; 66:363–376. [PubMed: 15283062]

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. [PubMed: 9577239]

Joanisee MF, Robertson EK, Newman RL. Mismatch negativity reflects sensory and phonetic speech processing. NeuroReport. 2007; 18:901–905. [PubMed: 17515798]

Klatt DH. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America. 1980; 67:971–995.

Kuhl PK, Miller JD. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. Science. 1975; 190:69–72. [PubMed: 1166301]

Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology. 1957; 54:358–368. [PubMed: 13481283]

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]

Maiste AC, Wiens AS, Hunt MJ, Scherg M, Picton TW. Event-related potentials and the categorical perception of speech sounds. Ear and Hearing. 1995; 16:68–89. [PubMed: 7774771]

Massaro DW, Cohen MM. Categorical or continuous speech perception: A new test. Speech Communication. 1983; 2:15–35.

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18:1–86. [PubMed: 3753912]

McMurray, B. KlattWorks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research. 2009. Manuscript in preparation

McMurray B, Aslin RN, Tanenhaus MK, Spivey M, Subik D. Gradient sensitivity to within-category variation in speech: Implications for categorical perception. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34:1609–1631. [PubMed: 19045996]

McMurray B, Samelson V, Lee S, Tomblin JB. Eye-movements reveal the time-course of online spoken word recognition language impaired and normal adolescents. Cognitive Psychology. 2010; 60:1–39. [PubMed: 19836014]

McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. Cognition. 2002; 86:B33–B42. [PubMed: 12435537]

McMurray B, Tanenhaus MK, Aslin RN. Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. Journal of Memory and Language. 2009; 60:65–91. [PubMed: 20046217]

Miller JL. Internal structure of phonetic categories. Language and Cognitive Processes. 1997; 12:865–870.

Näätänen R, Picton TW. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology. 1987; 24:375–425. [PubMed: 3615753]

Oden GC, Massaro DW. Integration of feature information in speech perception. Psychological Review. 1978; 85:172–191. [PubMed: 663005]

Ohlemiller KK, Jones LB, Heidbreder AF, Clark WW, Miller JD. Voicing judgments by chinchillas trained with a reward paradigm. Behavioural Brain Research. 1999; 100:185–195. [PubMed: 10212066]

Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, et al. Auditory cortex accesses phonological categories: An MEG mismatch study. Journal of Cognitive Neuroscience. 2000; 12:1038–1055. [PubMed: 11177423]

Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics. 1973; 13:253–260. [PubMed: 23226880]

Pisoni DB, Tash J. Reaction times to comparisons within and across phonetic categories. Perception and Psychophysics. 1974; 15:285–290. [PubMed: 23226881]

Pratt, H. Sensory ERP Components. In: Luck, SJ.; Kappenman, ES., editors. Oxford Handbook of Event-Related Potential Components. Oxford University Press; New York: (in press)

Repp, BH. Categorical perception: Issues, methods and findings. In: Lass, N., editor. Speech and Language: Advances in Basic Research and Practice. Academic Press; New York: 1984. p. 244-335.

Sams M, Aulanko R, Aaltonen O, Näätänen R. Event-related potentials to infrequent changes in synthesized phonetic stimuli. Journal of Cognitive Neuroscience. 1990; 2:344–357.

Schouten E, Gerrits B, van Hessen A. The end of categorical perception as we know it. Speech Communication. 2003; 41:71–80.

Sharma A, Dorman MF. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. Journal of the Acoustical Society of America. 1999; 106:1078–1083. [PubMed: 10462812]

Sharma A, Kraus N, McGee T, Carrell T, Nicol T. Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. Electroencephalography and Clinical Neurophysiology. 1993; 88:64–71. [PubMed: 7681392]

Sharma A, Marsh CM, Dorman MF. Relationship between N1 evoked potential morphology and the perception of voicing. Journal of the Acoustical Society of America. 2000; 108:3030–3035. [PubMed: 11144595]

Sinex DG, McDonald LP, Mott JB. Neural correlates of nonmonotonic temporal acuity for voice onset time. Journal of the Acoustical Society of America. 1991; 90:2441–2449. [PubMed: 1774413]

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. Journal of Neurophysiology. 1999; 82:2346–2357. [PubMed: 10561410]

Thibodeau LM, Sussman HM. Performance on a test of categorical perception of speech in normal and communication disordered children. Journal of Phonetics. 1979; 7:375–391.

Werker JF, Tees RC. Speech perception in severely disabled and average reading children. Canadian Journal of Psychology. 1987; 41:48–61. [PubMed: 3502888]

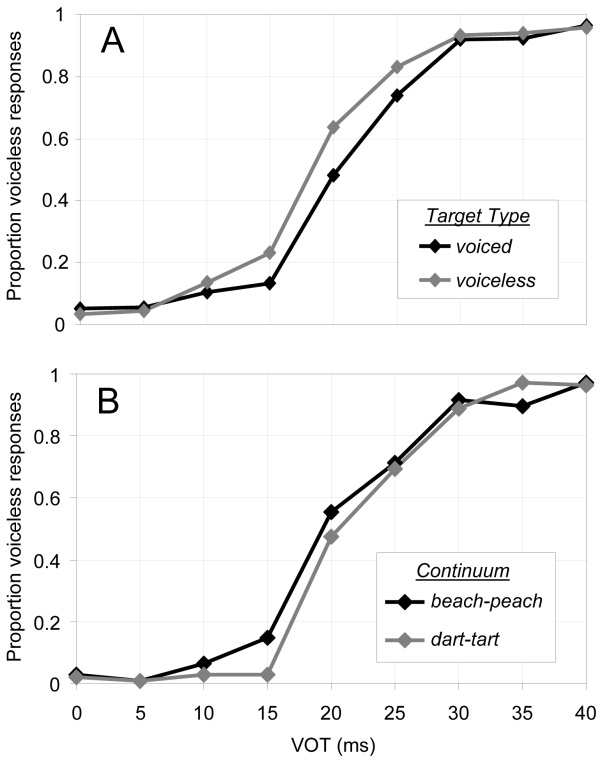

Fig. 1.

(A) Proportion of voiceless responses during the ERP task as a function of the nine VOT conditions for each of the two target voicing conditions. (B) Proportion of voiceless responses during the categorization task after the ERP recording session for each of the nine VOT conditions and two stimulus continua.

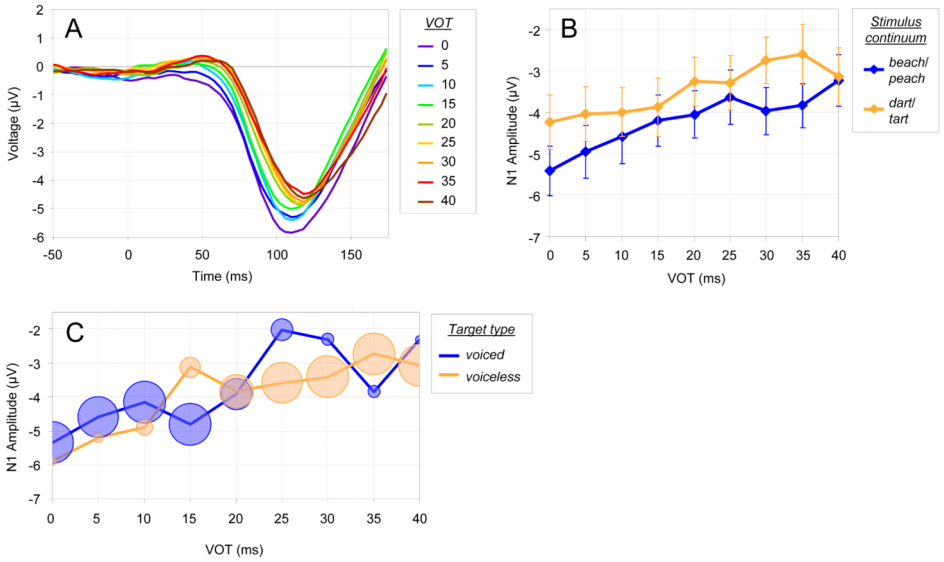

Fig. 2.

N1 results. (A) Grand average ERP waveforms, averaged across frontal electrodes, for each VOT condition. (B) Mean N1 amplitude as a function of the nine VOT conditions and two stimulus continuum conditions (beach/peach and dart/tart). Error bars represent standard error. (C) Mean N1 amplitude as a function of VOT and target voicing (voiced or voiceless) for trials where participants made target responses. The size of each data point in the figure is proportional to the number of trials for that condition.

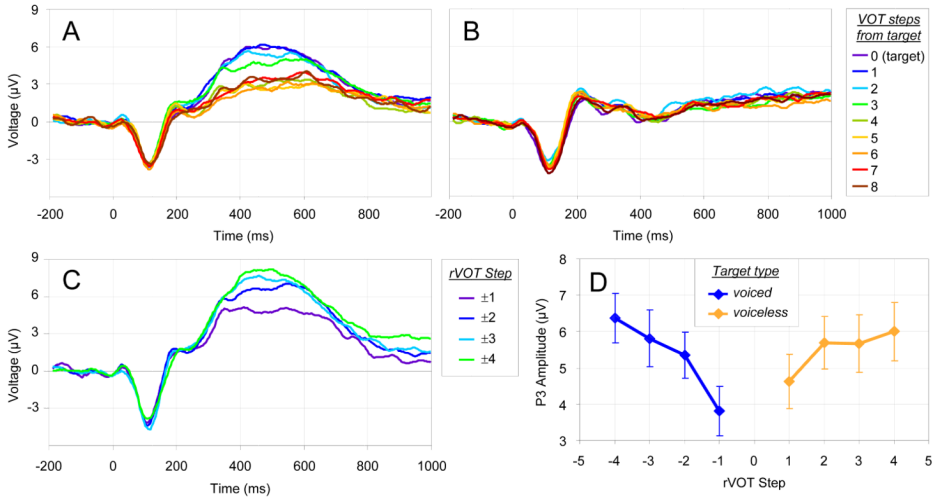

Fig. 3.

P3 results. (A) Grand average ERP waveforms, averaged across parietal electrodes, for each distance (in 5 ms VOT steps) from the target endpoint VOT (i.e. 0 ms if the target was voiced [beach or dart] or 40 ms if the target was voiceless [peach or tart]) for the relevant stimulus continuum. (B) Same as A, but for the irrelevant continuum. (C) Grand average ERP waveforms for target-response trials, relative to each participant's category boundary (rVOT steps; negative for voiced, positive for voiceless) rounded away from zero. (D) Mean P3 amplitude as a function of eight rVOT steps for target-response trials. Error bars represent standard error.