



## Research Article

## Individual differences in categorical perception of speech: Cue weighting and executive function

Eun Jong Kong<sup>a,\*</sup>, Jan Edwards<sup>b</sup><sup>a</sup> Korea Aerospace University, 100, Hanggongdae gil, Hwajeon-dong, Deogyang-gu, Goyang-city, Gyeonggi-do 412-791, South Korea<sup>b</sup> University of Wisconsin-Madison, 301 Goodnight Hall, 1975 Willow Dr., Madison, WI 53706, USA

## ARTICLE INFO

*Article history:*

Received 29 January 2015

Received in revised form

14 August 2016

Accepted 29 August 2016

Available online 23 September 2016

*Keywords:*

Individual difference

Categorical perception

Visual analogue scaling

Eye movement

Stop voicing contrast

Executive function capacity

## ABSTRACT

This study examined individual differences in categorical perception and the use of multiple acoustic cues in the perception of the stop voicing contrast. Goals were to investigate whether gradiency of speech perception was related to listeners' differential sensitivity to acoustic cues and to individual differences in executive function. The experiment included two speech perception tasks (visual analogue scaling [VAS] and anticipatory eye movement [AEM]) administered to 30 English-speaking adults in two separate experimental sessions. Stimuli were a /ta/ to /da/ continuum that systematically varied VOT and *f*0. Findings were that some listeners had a more gradient pattern of responses on the VAS task; the listeners who had a gradient response pattern on the VAS task also showed more sensitivity to *f*0 on the AEM task. The patterns were consistent across individuals tested on two separate occasions. These results suggest that variability in how categorically listeners perceive speech sounds is consistent and systematic within individuals.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Categorical perception of speech is a foundational principle of our understanding of how individuals process speech. Early research on speech perception observed that listeners perceive speech sounds categorically. That is, there is both a steep identification function in the perception of two speech sounds that are minimally different and a peak in the discrimination function between these two sounds that is centered on this category boundary (e.g., Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Harris, Kinney, & Lane, 1961). This finding was interpreted as support for the claim that listeners discard irrelevant subphonemic fine phonetic detail and pay attention only to higher-level categorical phonetic information. That is, categorical perception is important because it provides an explanation for how speech perception could be fast and accurate in the face of extreme variability in the acoustic signal, both within and across listeners.

Subsequent research found that listeners attend to – rather than discard – lower-level phonetic detail when processing speech, at least in some conditions. Talker differences, for example, are lexically irrelevant, but listeners consistently perform better in a single-speaker condition as compared to a multiple-speaker condition across a variety of experimental paradigms (e.g., Goldinger, Pisoni, & Logan, 1991; Luce & Lyons 1998; Mullenix, Pisoni, & Martin 1989, among many others). Furthermore, categorical perception is highly task-dependent; listeners perceive speech more categorically in some tasks than in others (e.g., Camey, Widen, & Viemeister, 1977; Schouten, Gerrits, & van Hessen, 2003). Massaro and Cohen (1983) used a Visual Analogue Scaling (VAS) task, in which listeners were given more options to respond continuously than the traditional two-alternative forced choice (2AFC) paradigm and found that listeners' responses were much less categorical. Massaro and Cohen found that the observed distribution of perceptual ratings on the VAS task was better fit with a model that assumed a continuous distribution than with a model that assumed a binary (discrete) distribution with individual listener variation. That is, listeners, as a group, showed *gradient* rather than categorical perception of stimuli on a task (VAS) that asked for a gradient response and categorical rather than gradient perception on a task (2AFC) that asked for a categorical response. However, it should be noted that even on

\* Corresponding author. Fax: +82 2 300 0493.

E-mail addresses: [ekong@kau.ac.kr](mailto:ekong@kau.ac.kr) (E.J. Kong), [jedwards2@wisc.edu](mailto:jedwards2@wisc.edu) (J. Edwards).

2AFC tasks, listeners tend to have longer reaction times for stimuli near a category boundary, (Pisoni & Tash, 1974; Studdert-Kennedy, Liberman, & Stevens, 1963; Repp, 1981), suggesting that even the 2AFC task is somewhat sensitive to response gradiency.

The visual world paradigm has also provided evidence for gradient rather than categorical perception of speech (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; McMurray, Tanenhaus, Aslin, & Spivey, 2003; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2009). For example, McMurray et al. (2003) used this paradigm to show that the proportion of looks to a pictured object is sensitive to changes in within-category acoustic values. The current study also uses an eye-tracking paradigm to explore whether there are individual differences in how categorically listeners perceive speech sounds and, if so, to examine whether these individual differences are related to listeners' sensitivity to sub-phonemic details.

Because of the early emphasis in the field on categorical perception, there has been little research on individual differences in speech perception. If speech is perceived categorically and acoustically irrelevant detail is discarded, then differences among normal-hearing listeners with typical speech and language development should be minimal and relatively uninteresting. Given this reasoning, researchers focused on differences as a function of different groups or conditions and mostly ignored individual differences. This research has established that the group differences in categorical perception are observed consistently. Developmental studies have reported that children relative to adults, and children with language or reading disorders relative to typically developing children, have both shallower slopes and reduced endpoint values in the probability estimation of choosing one category over the other using a 2AFC paradigm (e.g., Hazan & Barrett, 2000; Joanisse, Manis, Keating & Seidenberg, 2000; Werker & Tees, 1987). A common interpretation of such findings is that evidence of categorical perception indicates a robust perceptual representation of the target phonological contrast.

The few studies that have examined these individual differences have investigated cue-weighting strategies in speech perception (e.g., Hazan & Rosen, 1991; Idemaru, Holt, & Seltman, 2012; Kong & Edwards, 2011). Contrasts between speech sounds are typically signalled by a number of redundant acoustic cues. For example, both voice onset time (VOT, the latency between the oral release and the vocal fold vibration, Lisker & Abramson, 1964) and fundamental frequency ( $f_0$ ) at voicing onset are cues to the voicing contrast in stop consonants (e.g., Haggard, Summerfield & Roberts, 1981; Kingston & Diehl, 1994; Ohde, 1984; Whalen, Abramson, Lisker, & Mody, 1993). Voiced stop consonants in English are associated with a short-lag VOT and a lower fundamental frequency ( $f_0$ ) at vowel onset while voiceless stops are characterized with a long-lag VOT and a higher  $f_0$ . Using a variety of experimental paradigms, many studies have shown that the VOT cue is primary and the  $f_0$  cue is secondary in the perception of the English voicing contrast (e.g., Abramson & Lisker, 1985; Gordon, Eberhardt, & Rueckl, 1993; Whalen et al., 1993; Holt & Lotto, 2006; Francis, Kaganovich, & Driscoll-Huber, 2008; Idemaru & Holt, 2011; Francis & Nusbaum, 2002). There is some evidence, however, that there are individual differences in how these two cues are weighted, at least in some experimental conditions (e.g., Haggard, Ambler, & Callow, 1970; Hazan & Rosen, 1991; Idemaru et al., 2012; Walley & Carrell, 1983). Haggard et al. (1970) reported that some listeners were more sensitive to  $f_0$  than to VOT in an identification task of voiced and voiceless stop consonants. Stevens and Klatt (1974) also documented that some listeners relied more on VOT, while other listeners relied more on F1 onset frequency (another cue to the voicing contrast) to differentiate voiced and voiceless stops. More recently, Shultz, Francis, and Llanos (2012) found that English speakers differed in the extent to which they relied on the secondary cue of  $f_0$  to differentiate /b/ and /p/. Hazan and Rosen (1991) was one of the few studies to focus specifically on individual differences in cue weighting and categorical perception. They found that averaging across participants obscured large individual variability present in the identification curves for a place contrast (/b/-d/ and /d/-g/) between full-cue and reduced-cue conditions. They observed that listeners' identification functions were uniformly steep in the full-cue condition, whereas more individual variability was observed in the categorization slopes in the reduced-cue condition where less redundancy of the acoustic cues was provided. Francis et al. (2008) and Gordon et al. (1993) also observed individual differences in cue weighting in more demanding listening conditions. Both papers propose that listeners tend to rely on primary cues (such as VOT for the English stop voicing contrast) in ideal listening conditions, but adjust their cue-weighting toward secondary cues under less-than-ideal conditions, such as listening to speech in noise, or listening to multiple speakers, or listening when attentional resources are limited. These results are compatible with the concept of *dynamic redistribution of attention* in difficult listening conditions, as proposed by Francis et al. (2008).

In our own previous research (Kong & Edwards, 2011), we also observed individual differences in how categorically a voicing contrast was perceived as well as a difference in cue-weighting strategies. We found that some listeners had a gradient response pattern while others had a categorical response pattern when asked to differentiate between a voiced and voiceless stop consonant using a VAS task. Using an anticipatory eye movement paradigm (AEM, McMurray et al., 2003), we found that the same listeners who had a more gradient response pattern on the VAS task were much more sensitive to the secondary cue of  $f_0$ . That is, listeners who were more sensitive to  $f_0$  on the AEM task were also more likely to use the entire scale from /d/ to /h/ to characterize stimuli that differed in both VOT and  $f_0$ . By contrast, individuals who were less sensitive to  $f_0$  on the AEM task had a more categorical response pattern on the VAS task; they were more likely to rely primarily on the endpoints of the scale to characterize the same set of stimuli. These findings were important because they provided evidence that individual differences in cue-weighting of a secondary cue ( $f_0$ ) were associated with how categorically listeners perceived the same contrast in an off-line task (VAS). Still, the study was methodologically limited in the conclusions that it could draw for two reasons. First, listeners were tested only once, so it was impossible to determine whether the individual differences that were observed were related to differences in strategies that listeners might choose during a particular session or to consistent individual differences across participants. Furthermore, differences between categorical and gradient patterns of VAS responses were determined by visual inspection alone and listeners who were not clearly categorical or clearly gradient were excluded from the analysis. The current study was designed to address the limitations of the previous study by testing listeners in two separate sessions and by developing a measure to quantify the "gradiency" of their response patterns on the VAS task.

A secondary goal of this study was to explore whether individual differences in gradiency of response on the VAS task might be related to individual differences in executive function. Executive function refers to “a general-purpose control mechanism that modulates the operation of various cognitive subprocesses (Miyake et al., 2000; p.50)” such as inhibiting distraction, expanding working memory, planning, controlling attention, shifting between tasks, updating and monitoring information (Karpicke, Conway, & Pisoni, 2007; Festman, Rodriguez-Fornells, & Münte, 2010; Bialystok & Craik, 2010). While several studies have observed individual differences in patterns of cue-weighting, these studies have not addressed what might cause these differences (e.g., Hazan & Rosen, 1991; Kong & Edwards, 2011; Idemaru et al., 2012). We hypothesized that individual differences in individual differences in sensitivity to sub-phonemic details in speech perception might be related to individual differences in executive function. While the relationship between executive function and speech perception has not been studied extensively in young adults with normal hearing, it has been studied in populations with hearing impairment. For example, it is well-known that inhibitory processes decrease as a function of aging and it has been posited that one explanation for older listeners’ difficulty in understanding speech in noise is related to their problems with inhibiting the processing of irrelevant information – i.e., background noise (Gordon-Salant & Fitzgibbon, 2004). Humes (2002) found that digit span, a measure of working memory, was a significant predictor of speech recognition scores in older adults with hearing aids. Recent research by Pisoni and colleagues has focused on whether individual differences in executive function can explain individual differences in word and sentence recognition in children with cochlear implants and normal-hearing children listening to vocoded speech (which simulates the auditory degradation of a cochlear implant). Karpicke et al. (2007) found, in a study of normal hearing children aged 5–8 years, that individual differences in a measure of inhibition were significantly correlated with individual differences in a sentence recognition task with vocoded speech. Most relevant to the current study is a study by Weiss, Gerfen, and Mitchel (2010). Weiss and colleagues investigated the ability of young normal-hearing adults to attend to multiple cues (statistical and duration cues) to word segmentation. When both cues were of equal strength, individual differences in successful word segmentation were related to individual differences on the Simon task, a measure of inhibitory control. Given the findings of Weiss et al. (2010) as well as previous studies with different populations, we wondered if individual differences in the gradiency of response on a VAS task and perceptual weighting of a secondary cue to voicing ( $f_0$ ) might be related to individual differences in executive function capacity. Such a finding would provide further support for the claim that the individual differences in sensitivity to  $f_0$  observed by Kong and Edwards (2011) are consistent properties of individual listeners. More generally, the finding would provide experimental evidence that listeners’ speech processing strategies are modulated by their executive function capacity.

To summarize, the purpose of this paper was to examine whether individual differences in processing speech categories were consistent and systematic within a speaker. We were interested in confirming the observation of an earlier small-scale study by Kong and Edwards (2011) that listeners who had a more gradient response pattern on a visual analogue scaling task were more likely to utilize a secondary cue in online processing (the AEM task). This study expands on the previous study in several respects: we tested a larger group of participants two times with the same AEM and VAS tasks and we developed a method to quantify the gradiency of participants’ response patterns on the VAS task. We also investigated whether individual differences in processing speech (gradiency of response and weighting of  $f_0$ , a secondary cue) were related to individual differences in executive function capacity.

## 2. Perception experiments

Two speech perception experiments (VAS and AEM) were conducted to explore whether there were individual differences in gradiency (or categoricity) of response on a VAS task and, if so, whether these differences were related to listeners’ differential sensitivity to  $f_0$  on two perception experiments. Both the VAS and AEM tasks were administered twice, separated by about a week to evaluate whether listeners showed consistent response patterns over two test sessions. In addition to the speech perception experiments, all listeners were evaluated on several measures of executive function and a measure of non-verbal IQ.

### 2.1. Participants

The participants were 30 native English-speaking adults with no self-reported history of speech, language, or hearing problems (24 females, 6 males, age range: 18 to 24 years). All participants were undergraduate students at the University of Wisconsin-Madison and either received course credit or were compensated for their participation.

### 2.2. Stimuli

The stimuli were pseudo-synthetic consonant-vowel (CV) syllables that were constructed to make a continuum from /ta/ to /da/. They were synthesized using words (*tot* and *dot*) produced by an adult male speaker from Wisconsin. We selected one token of /da/ and we systematically varied both VOT and  $f_0$  to create the continuum. VOT values were manipulated by excising a portion of the burst release/aspiration from /ta/ and pasting it before the voicing onset of the /da/ token. Six different log-scale steps of VOT were included: 9 ms (original VOT of /da/), 13 ms, 19 ms, 28 ms, 40 ms, and 59 ms. Log-scale steps were chosen because of the non-linear nature of perceptual sensitivity to changes in VOT values. At each VOT step, we replaced the original  $f_0$  value during the vowel (165 ms duration) with one of five different sustained  $f_0$  values (98 Hz, 106 Hz, 114 Hz, 122 Hz, and 130 Hz: Whalen et al., 1993) using Praat (Boersma & Weenink, 2014). A total of 30 different stimuli (a 6-step VOT  $\times$  5-step  $f_0$  continuum) were created. CV syllables rather than words were used as stimuli in both

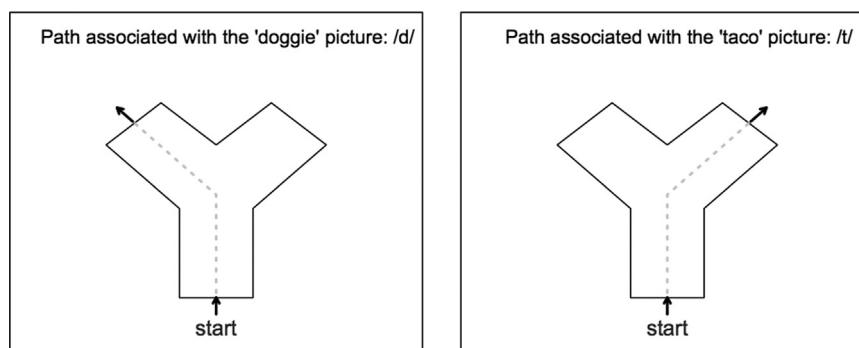


Fig. 1. Illustration of the Y-shape pipe used in the anticipatory eye movement task.

experimental tasks for two reasons: one, we wanted to minimize any lexical confounds such as word frequency; and two, we wanted the same stimuli for both tasks and the VAS task did not include visual prompts.

### 2.3. Tasks

There were two experimental tasks, a visual analogue scaling task (VAS) and an anticipatory eye movement task (AEM). In a VAS rating task, individuals are asked to scale a psychophysical parameter by indicating their percept on an idealized visual display. In our VAS task, an arrow was displayed on the computer monitor immediately after each stimulus was played. One end of the arrow was labeled as the 'd' sound and the other end of the arrow was labeled as the 't' sound. Listeners were instructed to click anywhere on the arrow, based on their judgment of how close the stimulus was to either /da/ or /ta/. Listeners heard the same 30 stimuli items three times in random order. It should be noted that the number of stimuli items on the VAS task presented in each experimental session was relatively low (90 items), relative to many speech perception experiments (e.g., Liberman et al., 1961; Pisoni & Tash, 1974; Massaro & Cohen, 1983).

In the AEM task, we used the SMART-T program to implement an anticipatory eye movement paradigm using the Tobii 2150 eye-tracker (Shukla, Wen, White, & Aslin, 2011). For all trials including six training trials, when an auditory stimulus item was played, a picture (representing either *doggie* or *taco*, words which contain a syllable-initial /da/ or /ta/ in the local dialect) appeared at the bottom of the pipe and moved slowly through the pipe. The /da/ stimuli were paired with the picture of *doggie* and consistently appeared on the left side of the Y-shaped pipe (Fig. 1, left panel), while the /ta/ stimuli were paired with the picture of *taco* and consistently appeared on the right side of the pipe (Fig. 1, right panel).<sup>1</sup> Listeners were conditioned to make anticipatory looks to either the left or the right side of a Y-shaped pipe based on whether the sound they heard was more similar to /d/ or /t/. The Y-shaped pipe was transparent at the beginning of the training trials so that participants could see the path of the moving picture, and gradually the pipe became opaque during the six training trials. During the experimental trials, the pipe was opaque and the participants had to anticipate on which side the picture would appear based on whether they perceived the auditory stimulus to be /d/ or /t/. The participants were given no other instructions beyond "look at the computer screen." Each stimulus was presented four times in the AEM task. Ambiguous stimulus items (those with VOT values of 19 ms or 28 ms) appeared twice on the left side of the pipe and twice on the right side in a random order to prevent listeners from responding based on local statistics. While the AEM paradigm has been used extensively in infant research, it has not been used in many adult studies (although Richardson and Kirkham (2004) included an adult control group). We chose the AEM paradigm over the more frequently-used visual world paradigm because the AEM paradigm is ideal for eliciting judgments of multiple presentations of speech stimuli without worrying about lexical confounds.

Listeners participated in two experimental sessions separated by about a week. In the first session, the two tasks of VAS and AEM were counter-balanced, so approximately 50% of the participants did the AEM experiment first and the VAS experiment second, while the other 50% had the reverse order in their first session. In the second experimental session, about half of the returning participants were given the two experiments in the same order as the first session, while the other half were given the experiments in the opposite order.

We also administered several standardized tests during the second test session after the participants had completed the two speech perception tasks. Two subtests from the DKEFS (*Delis-Kaplan Executive Function System*, Delis, Kaplan, & Kramer, 2001) were administered to measure components of executive function: the Trail-Making Task (TMT), and the Color-Word Naming (CW) subtests. The Trail-Making task is a paper-and-pencil task with three conditions. Participants are asked to connect number sequences (e.g., 1 to 2 to 3) in Condition 1, letter sequences (e.g., A to B to C) in Condition 2, and number-letter sequences (e.g., 1 to A to 2 to B, etc.) in Condition 3 as quickly as possible. In the Color-Word Naming task, participants are asked to name the font color of a word in three conditions: (1) congruent color-word naming (e.g., the word "red" in a red font), (2) incongruent color-word naming (e.g., the word "red" in a green font), and (3) switch color-word naming (some trials are congruent and some are incongruent). The

<sup>1</sup> It should be noted that the fact that the /d/-initial word always appeared on the right side of the pipe and the /t/-initial word on the left side across listeners is a design limitation. While this is an experimental confound, it is unlikely to have influenced the results (as it is not very likely that there would have been a consistent looking preference for short VOTs to one side of the screen). A counter-balanced design in which half of the participants received /d/ on the left side and half received /d/ on the right side would have been preferable.

incongruent and switch conditions of the Color-Word Naming subtests are considered measures of inhibition while Condition 3 of the Trail-Making Task are considered measures of cognitive flexibility or attention switching (Miyake et al., 2000). These subtests were given in a randomized order. Duration and accuracy were obtained for each condition for all two tasks.

Non-verbal IQ was assessed using the matrices section of the Kaufman Brief Intelligence Test – 2nd edition (KBIT-2, Kaufman & Kaufman, 2004). This test was given so that we could partial out any general effect of intelligence that might be correlated with executive function (e.g., Friedman et al., 2006; Festman et al., 2010).

## 2.4. Analysis

### 2.4.1. VAS task

For the VAS task, we examined the response patterns of individual listeners. We were interested in two questions: one, were individual listeners more categorical or gradient in their responses patterns; and two, how did individual listeners attend to the different combinations of VOT and  $f_0$ .

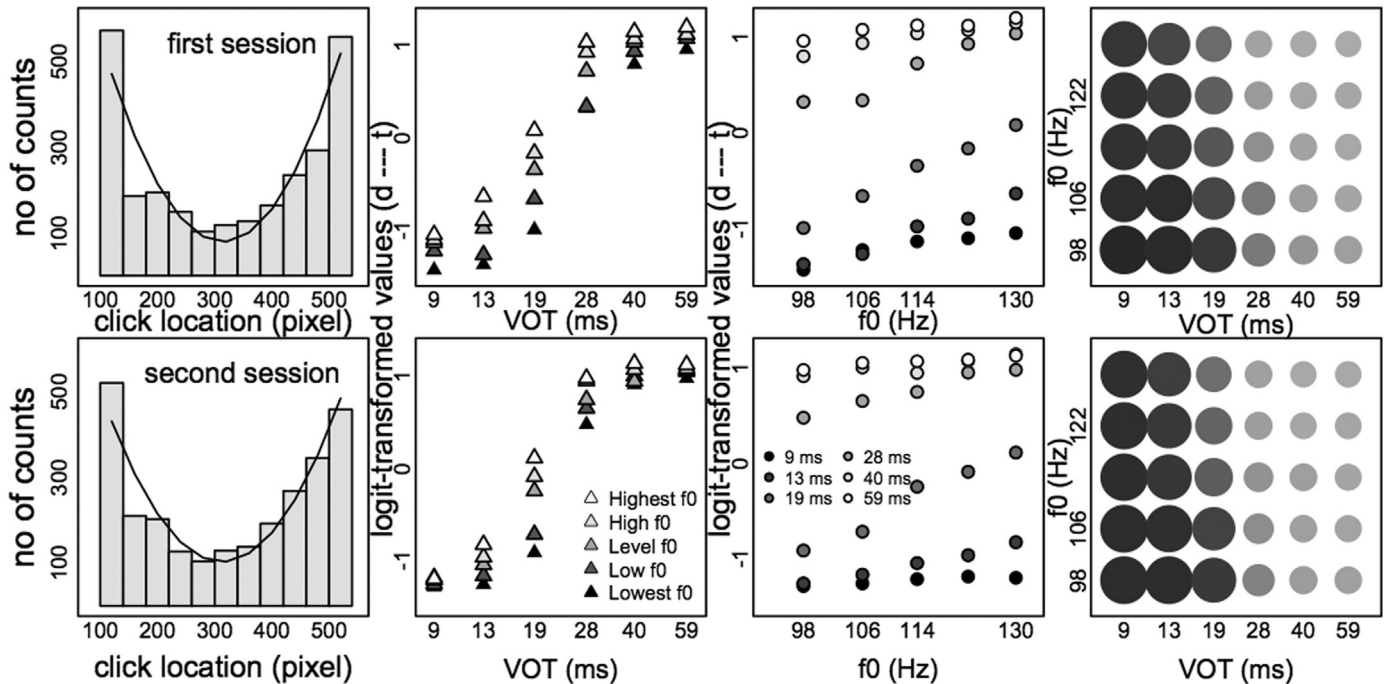
First, we examined individual differences in gradiency of response by comparing histograms of click locations (in pixels) for each individual listener. In Kong and Edwards (2011), we had observed non-uniform histogram distributions across subjects; response patterns by some listeners resulted in bipolar peaks at the two ends of the arrow, while response patterns for other listeners resulted in click distributions that were more evenly spread across the entire arrow. To quantify these differences in shape, we constructed a polynomial function fitting the histograms, implemented in R (R Core Team, 2016), and used the coefficients of the quadratic terms as a numerical index of *gradiency* (or categoricity) for each participant. That is, we interpreted larger coefficients from the concave curve shapes as indicating a more categorical response pattern on the VAS task. Conversely, we interpreted smaller coefficients from these curves as indicating a more gradient response pattern on this task. This numerical index could capture varying degrees of spread of the click distributions for most participants (see the response pattern averaged across the entire group in Fig. 2 and three individual response patterns in Fig. 3). We used this measure of gradiency for four purposes: one, to evaluate whether individual response patterns were similar across the two sessions; two, to evaluate whether gradiency of response on the VAS task was related to sensitivity to VOT or  $f_0$  on the VAS task; three, to evaluate whether gradiency of response was related to sensitivity to  $f_0$  on the AEM task; and four, to examine the relationship between gradiency on the VAS task and performance on the two measures of executive function.

In relating the responses to the acoustics of the stimuli, we constructed mixed-effects regression models to examine how changes in VOT and  $f_0$  values of the stimuli affected listeners' perception of the /t-/d/ contrast. The pixel values of the click locations were the dependent variable after being logit-transformed due to the bounded ends of the pixel data. VOT,  $f_0$  and the interaction between  $f_0$  and VOT were the fixed effect variables in these models. The responses of individual listeners were considered to be a random effect, so the slopes of VOT and  $f_0$  were allowed to vary. The regression models were implemented using the *lme4* package in R (Bates, Maechler, & Bolker, 2011). Since the *lmer* command in the *lme4* package does not provide *p*-values, we either computed *p*-values based on the parameter estimates and the standard errors by transforming the *t*-distribution into a *z*-distribution or we conducted deviance tests where the full model was compared to a simpler model in order to estimate the effect of a selectively-added or removed factor.

### 2.4.2. AEM task

The eye gaze data from the AEM task were dichotomously coded: looks to /d/ were coded as '1' and looks elsewhere were coded as '0'. We grouped these discrete observations into a series of temporal bins (50 ms bins) within each listener's trials of the same stimuli and calculated the empirical logit of looks to /d/ in each bin as described in Barr (2008). For each VOT condition, we constructed a mixed-effects model with the estimated logit value of looking to /d/ as the dependent measure and time (temporal bins) as a fixed effect variable, and  $f_0$  as the interaction term. Similarly, for each  $f_0$  condition, we constructed a mixed-effects model with VOT as the interaction term. This type of time-series regression has been proposed as a way to model how looks to a target change over time for the visual world paradigm (e.g., Mirman, Dixon, & Magnuson, 2008; Barr, 2008; Singer & Willett, 2003). To our knowledge, this is the first use of growth curve analysis to model eye-tracking responses from an AEM experimental paradigm. We used an analysis window that began at 200 ms after the auditory stimulus onset (because it takes about 200 ms to program an eye movement, Matin, Shao, & Boff, 1993) and that ended at the onset of the visual stimulus (the picture of *doggie* or *taco* which appeared on the left or right side of the Y-shaped pipe). The model included three time terms (first order or linear, second order or quadratic, and third order or cubic) using an orthogonal polynomial coding scheme. As described in Mirman et al. (2008), we transformed these time terms so that they were orthogonal to each other in order to eliminate correlations among them. Another advantage of using orthogonal time is that the intercept term is not fixed to a particular temporal bin, but can be interpreted as corresponding to the average shape of the growth curve. The slope of the linear time term is most relevant to our assessment of listeners' sensitivity to changes in VOT and  $f_0$ . Larger absolute values of the first order coefficients indicate that linear slopes are steeper (i.e., faster looks to the /d/ side of the screen). In addition, the intercept term and the second and third order time terms provide information on the shape of the curve independently of the linear term (or slope). Briefly, the intercept describes the overall height of the curve, the quadratic component describes the symmetrical rise (or fall) rate of the curve, and the cubic component describes the asymmetric rise (or fall) rate near the trough (or peak) of the curve (Mirman et al., 2008; Singer & Willett, 2003). The mixed-effect models included random effects of subject and subject-by-stimulus condition, where the intercepts and slope coefficients of the first polynomial time term were allowed to vary. Random effects of the quadratic and cubic components were not included so that we could ascribe random variation solely to the first order time term, the explanatory variable related to the experimental question.

In order to quantify sensitivity to  $f_0$  for individual listeners, we made two sets of comparisons of specific stimulus conditions in the mixed-effects regression models. First, we compared two sets of conflicting vs. cooperating cue conditions. These were: (1) the



**Fig. 2.** The leftmost column shows histograms of click locations of all listeners' responses with regression curves of quadratic terms overlaid on top. The two center columns show the logit-transformed click-location values averaged across listeners as a function of VOT (second column) and  $f_0$  (third column). The rightmost column shows all subjects' responses to each stimulus category in a VOT  $\times$   $f_0$  space; the darkness/size of the symbol is a function of how /d/-like the stimulus was rated (the larger and the darker the dot, the more /d/-like). The responses from the first session are presented in the top panels and those from the second in the bottom panels.

highest  $f_0$  condition (130 Hz  $f_0$ ) with either the shortest lag VOT (9 ms, a conflicting cue condition) or longest lag VOT (59 ms, a cooperating cue condition) VOT; and (2) the lowest  $f_0$  condition (98 Hz  $f_0$ ) with either the shortest lag VOT (9 ms, a cooperating cue condition) or longest lag VOT (59 ms, a conflicting cue condition). We also examined two ambiguous VOT conditions (19 ms and 28 ms) with either the lowest (98 Hz) or highest (130 Hz)  $f_0$  conditions.

We reasoned that comparing listeners' performance in a conflicting  $f_0$  cue condition (shortest lag VOT and highest  $f_0$  or longest lag VOT and lowest  $f_0$ ) to a cooperating  $f_0$  cue condition (longest lag VOT and highest  $f_0$  or shortest lag VOT and lowest  $f_0$ ) would best capture individual listeners' sensitivity to  $f_0$ . That is, in the competing cue conditions, listeners who were less sensitive to  $f_0$  would continue to attend primarily to VOT, even if the VOT information and the  $f_0$  information were in conflict. For these listeners, we expected that there would continue to be steep linear slopes in the shortest lag VOT condition, yielding large linear slope differences between the two VOT conditions (9 ms VOT and 59 ms VOT). By contrast, we hypothesized that listeners who were more sensitive to  $f_0$  would attend to  $f_0$  information in the conflicting cue conditions. Therefore, these listeners would have relatively less steep linear slopes in the shortest lag VOT condition, producing smaller linear slope differences between the two VOT conditions. To test this hypothesis, we subtracted the random effects coefficient of the 59 ms VOT condition (estimated at the subject-by-condition level) from that of the 9 ms VOT condition in the highest  $f_0$  condition. Similarly, we subtracted the random effect coefficient of the 9 ms VOT condition from that of the 59 ms VOT condition in the lowest  $f_0$  condition.

We also analyzed two ambiguous VOT conditions (19 ms and 28 ms) in a similar manner in order to examine individual listeners' sensitivity to  $f_0$ , as previous studies have shown that a redundant secondary cue plays a more important role when the primary cue is ambiguous (Abramson & Lisker, 1985; Whalen et al., 1993). We expected that listeners who were more sensitive to  $f_0$  in general would be better able to take advantage of this secondary cue in the ambiguous VOT conditions. That is, we predicted that the listeners who were more sensitive to  $f_0$  would have relatively greater linear slope differences between the high (130 Hz) and low (98 Hz)  $f_0$  conditions in the two ambiguous VOT conditions. By contrast, we predicted that the listeners who were less sensitive to  $f_0$  would have relatively smaller linear slope differences between these two conditions. To calculate these slope differences for individual listeners, we subtracted the random effects coefficient of the 130 Hz  $f_0$  condition (estimated at the subject-by-condition level) from that of the 98 Hz  $f_0$  condition.

Based on the results of Kong and Edwards (2011), we hypothesized that listeners who were more categorical on the VAS task would rely less on the secondary cue of  $f_0$ , whereas the listeners who were more gradient on this task would be more sensitive to the  $f_0$  cue. To test this hypothesis, we examined whether these measures of individuals' sensitivity to  $f_0$  in the AEM task – the difference in linear slope coefficients between the two VOT conditions in the high  $f_0$  condition difference in linear slope coefficients between the two  $f_0$  conditions in the ambiguous VOT conditions – were correlated with our gradiency measure from the VAS task.

#### 2.4.3. Cognitive measures in relation to gradiency of speech perception and sensitivity to $f_0$

We examined whether our measure of gradiency was related to measures of executive function and non-verbal IQ using partial correlations. For the two executive function subtests, we used response time to complete the task as the raw score and then obtained standard scores from the DKEFS manual. (Differences in accuracy across participants were minimal.) We used standard scores from

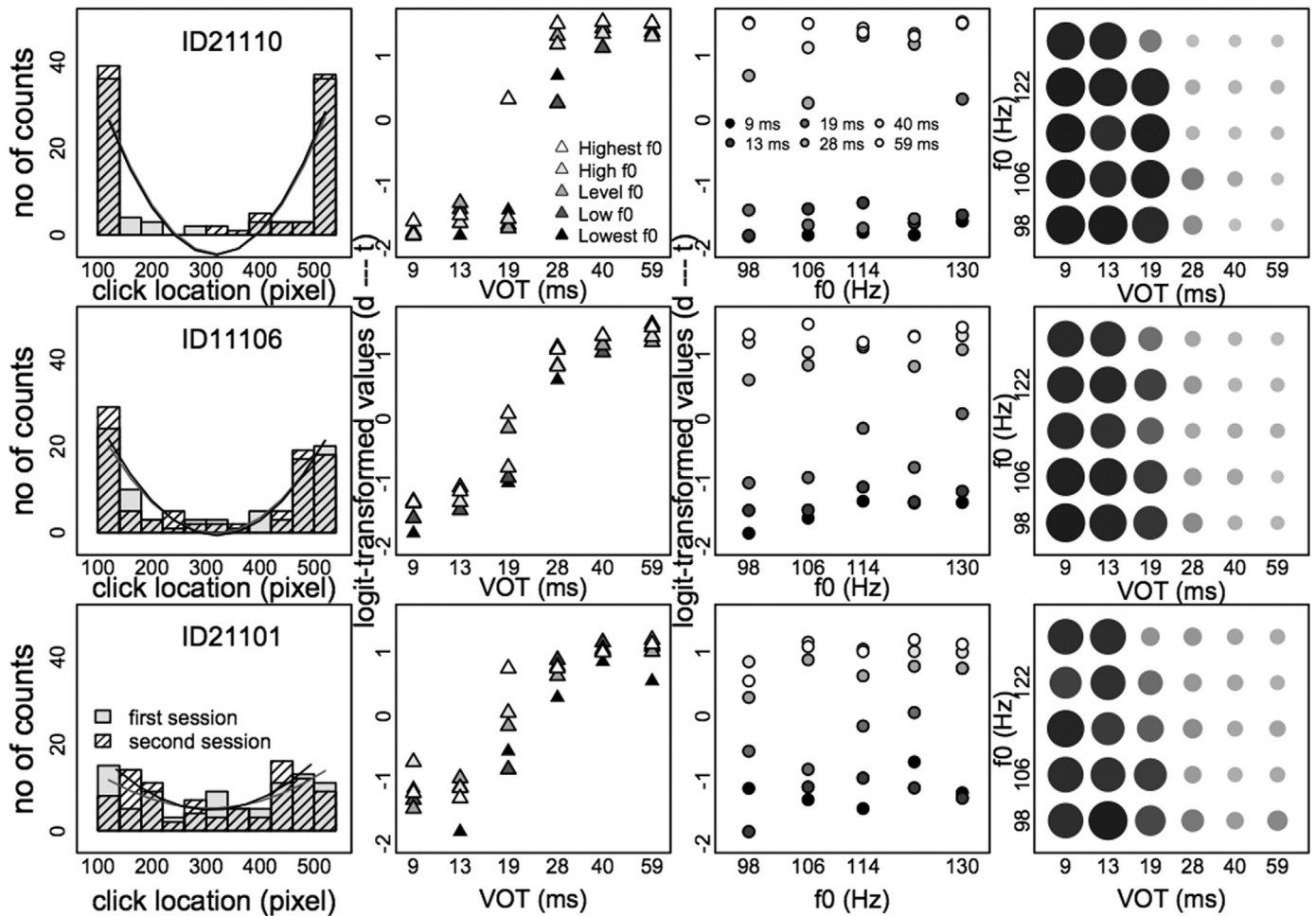


Fig. 3. Examples of three different response patterns for the VAS task: categorical (top), intermediate (middle), and gradient (bottom) from the first session. The leftmost column plots histograms of click locations with the quadratic curves overlaid based on the regression models. The logit-transformed click-location values in pixels averaged within subject are shown as a function of VOT (second column) and  $f_0$  (third column). The rightmost column plots individual subject responses to each stimulus category in a VOT  $\times$   $f_0$  space; the darkness/size of the symbol is a function of how d-like the stimulus was rated (the larger and the darker the dot, the more /d/-like).

the KBIT-2 as the measure of non-verbal IQ. To explore the relationship between gradiency and our measures of executive function, we completed two analyses. In the first analyses, we partialled out the effect of non-verbal IQ and then examined the relationship between our gradiency measure from the VAS task and each of the two executive function measures. In the second analysis, we partialled out both the effect of non-verbal IQ and one executive function measure and then examined the relationship between the gradiency measure and the other executive function measures. We also examined whether the two executive function measures were correlated with our measures of sensitivity to  $f_0$  from the AEM task (linear slope differences between models for the 9 ms and 59 ms VOT conditions in the highest  $f_0$  condition and linear slope differences between models for the 98 Hz and 130 Hz  $f_0$  conditions in the ambiguous VOT condition). Again, we performed two partial correlations, one in which only non-verbal IQ was partialled out and a second in which one of the executive function measures was also partialled out. The analyses were implemented using the *pcor.test* command of the *ppcor* package in R (Kim, 2012). We used a Bonferroni correction of the  $p$  value because of the multiple comparisons. Accordingly, the significance level was set at  $p < 0.0125$  which was equivalent to  $p < 0.05$  if running a single test.

### 3. Results

#### 3.1. VAS task

Fig. 2 shows the click distributions of the VAS task for all listeners (leftmost column) and the average ratings as a function of VOT and  $f_0$  (center columns). The responses from the first session are shown in the top panels and those from the second session are shown in the bottom panels.

The histogram in the top left panel exhibited two global peaks at each end of the continuum indicating listeners' overall tendency to perceive the stimuli categorically. However, it can be observed that there were click responses between the two peaks suggesting that at least some listeners were also sensitive to category goodness. The regression curves estimated by the quadratic term

regression model are overlaid on the histograms and schematically show the average response characteristics of the group in each of the two sessions.

The panels in the second and third columns of Fig. 2 plot the average rating values as a function of VOT and  $f_0$ , respectively. It can be observed that perception of the contrast between /t/ and /d/ was much more strongly influenced by VOT than by  $f_0$  in both sessions. The stimuli were more likely to be judged as more /t/-like as VOT values increased (second column), whereas the ratings did not change dramatically as  $f_0$  values changed (third column). Based on visual inspection, an effect of the  $f_0$  cue was observed primarily for stimuli in the ambiguous VOT (19 ms VOT and 28 ms) conditions. In these conditions, stimuli were judged as more /t/-like as  $f_0$  values increased. These observations were confirmed by the results of the mixed-effects regressions. The regression model included VOT,  $f_0$ , and the interaction of VOT and  $f_0$  as fixed effects and allowed the slopes of VOT and  $f_0$  to vary at the individual subject level. When this complex model was compared to a simpler regression model where the interaction of VOT and  $f_0$  was not included, the results of the deviance test (first session:  $\chi^2=13.7$ ,  $df=1$ ,  $p<0.001$ ; second session:  $\chi^2=6.03$ ,  $df=1$ ,  $p<0.05$ ), indicated that adding the interaction term significantly improved model fit for the data from both sessions. In the complex model where VOT,  $f_0$ , and the interaction of VOT and  $f_0$  were the explanatory variables and the measurement units were normalized, both VOT and  $f_0$  were significant predictors of listener response patterns in the two sessions. The effect of VOT was greater than that of  $f_0$ , as indicated by its larger coefficient in both sessions (first session:  $\beta_{\text{VOT}}=0.85$  [S.E.=0.048;  $t=17.5$ ] vs.  $\beta_{f_0}=0.20$  [S.E.=0.029;  $t=7.15$ ]; second session:  $\beta_{\text{VOT}}=0.85$  [S.E.=0.043;  $t=19.6$ ] vs.  $\beta_{f_0}=0.14$  [S.E.=0.017;  $t=8.3$ ]). This result confirmed listeners' greater sensitivity to VOT than to  $f_0$  in differentiating the contrast between /t/ and /d/ on the VAS task.

The listeners' responses to each stimulus item (that is, to each combination of VOT and  $f_0$ ) are shown in the rightmost column of Fig. 2, which was inspired by Escudero and Boersma (2004). The /d/-likeness of the listener's perception, based on the pixel numbers of actual clicks, is indicated by the size and the degree of darkness of each stimulus' symbol; the larger and darker the symbol, the more /d/-like the listener's perception of the stimulus. It can be observed that the symbols have darker colors for the shorter VOT values, indicating that subjects on average were strongly influenced by VOT in perceiving /d/ vs. /t/. This dark color is clearly in contrast with the lighter colors for the longer VOT values. By contrast, the size-and-darkness pattern was similar for the cells at the top of each panel (higher  $f_0$  stimuli) as compared to those at the bottom of the panel (lower  $f_0$  stimuli), indicating that the listeners, as a group, were not greatly influenced by  $f_0$ . The effect of  $f_0$  appeared to be conditioned by the VOT values of the stimuli. While the size-and-darkness pattern rarely changed within the same VOT conditions, the symbols gradually became smaller and lighter as  $f_0$  values increased in the mid-range of VOT values (19 ms and 28 ms), namely in the ambiguous VOT conditions.

These group results were consistent with previous research (e.g., Abramson & Lisker, 1985; Gordon et al., 1993; Whalen et al., 1993). However, we also observed that there were clear differences across participants in gradiency of response, as evidenced by differences in the histograms of click locations (leftmost column of Fig. 3). Fig. 2 shows histograms and regression curves estimated from each session of the VAS task for three selected participants who represented three different response patterns. Some listeners (e.g., ID21110 in the top panel) judged the stimuli categorically, choosing responses mostly at the two endpoints of /t/ and /d/. By contrast, other listeners (e.g., ID21101 in the bottom panel) judged the stimuli in a much more gradient manner, choosing responses across the entire VAS scale. There was also an intermediate pattern (neither clearly gradient nor clearly categorical) illustrated by the participant in the middle panel. The individual listeners' gradiency of response was quantified in terms of the slope coefficient of the quadratic term of the regression model (Section 2.4.1). As illustrated in the leftmost column panels of Fig. 3, the quadratic term estimated by the regression model yielded a relatively steep concave curve for the categorical listener ( $\beta[\text{quadratic term}]=0.00078$  for the first session; 0.0007 for the second) and a relatively shallow concave curve for the gradient listener ( $\beta[\text{quadratic term}]=0.00018$  for the first session; 0.00030 for the second). Importantly, these response patterns for individual listeners were consistent across the two test sessions, although the exact value of the gradiency measure varied from one session to another. A correlation test showed a significant positive correlation in individuals' gradiency of response (quadratic slope coefficients) between the two sessions ( $r=.48$ ,  $df=29$ ,  $p<.01$ ).

The other three columns in Fig. 3 display the VAS responses of the first session as a function of the two acoustic cues for these same three speakers. Similar to the group results, perception of the contrast between /t/ and /d/ was more strongly influenced by VOT than by  $f_0$  for all three of these listeners. The three plots in the second column of Fig. 3 (one for each speaker) show a strong relationship between VOT and the ratings; stimuli with longer VOT values were consistently rated as more /t/-like. In contrast, the three plots in the third column show that stimuli with higher  $f_0$  were not consistently rated as more /t/-like. Linear regressions with VOT,  $f_0$  and the interaction of the two as predictors, which were performed separately for the three subjects, indicated that VOT was a significant predictor ( $p<.0001$ ) of VAS responses for all three listeners in both VAS sessions but  $f_0$  was not consistently significant across listeners and the interaction of VOT and  $f_0$  was not significant across listeners in either session. The  $f_0$  coefficients for the responses of the participant shown in the top panel [ID21110] did not reach a significance level of  $p<.05$  in either session [first session:  $\beta_{f_0}=0.16$  ( $p=.074$ ); second session:  $\beta_{f_0}=0.13$  ( $p=.198$ )] and those for the responses of the participant shown in the middle panel [ID11106] were significant in only one session [first session:  $\beta_{f_0}=0.13$  ( $p=.072$ ); second session:  $\beta_{f_0}=0.20$  ( $p<.01$ )]. The  $f_0$  coefficients for the responses of the participant shown in the bottom panel [ID21101] were significant in both sessions [first session:  $\beta_{f_0}=0.16$  ( $p<.05$ ); second session:  $\beta_{f_0}=0.17$  ( $p<.005$ )]. When the measurement units were standardized, the regression models showed that VOT was a more effective variable than  $f_0$  in explaining the response variable; consistently for the three speakers, the coefficients of VOT were greater than those of  $f_0$ : First session: [ID21110]  $\beta_{\text{VOT}}=1.24$  vs.  $\beta_{f_0}=0.16$ ; [ID11106]  $\beta_{\text{VOT}}=1.18$  vs.  $\beta_{f_0}=0.13$ ; [ID21101]  $\beta_{\text{VOT}}=0.82$  vs.  $\beta_{f_0}=0.16$ ; second session: [ID21110]  $\beta_{\text{VOT}}=0.13$  vs.  $\beta_{f_0}=1.18$ ; [ID11106]  $\beta_{\text{VOT}}=1.07$  vs.  $\beta_{f_0}=0.20$ ; [ID21101]  $\beta_{\text{VOT}}=0.79$  vs.  $\beta_{f_0}=0.17$ .

The listeners' responses to each stimulus item in a VOT  $\times$   $f_0$  space are presented in the panels in the rightmost column of Fig. 3. As in Fig. 2, the /d/-likeness of the listener's perception is indicated by the size and the degree of darkness of each stimulus' symbol. It can be observed that the symbols in all three panels have darker colors for the shorter VOT values, indicating that all three subjects were strongly influenced by VOT in perceiving /d/ vs. /t/. The difference among these three listeners was that the dots for the categorical listener (top panel) become smaller and darker rather abruptly between 19 ms VOT and 28 ms VOT, whereas those of the gradient listener (bottom panel) become smaller and darker gradually as VOT values increase. If the size-and-darkness pattern is similar for the cells at the top of each panel (higher  $f_0$  stimuli) as compared to those at the bottom of the panel (lower  $f_0$  stimuli), then this indicates that the listener is not influenced by  $f_0$ . It can be observed that the size and darkness patterns of the symbols for the categorical listener in the top panel differed slightly from those of the gradient listener in the bottom panel. For the categorical listener, the size and darkness of the symbols varied very little between the top and bottom halves of the panel. By contrast, the plot for the gradient listener appeared to have some variation in the size and darkness of the symbols as the stimuli varied along the  $f_0$  dimension, suggesting that this listener was sensitive to  $f_0$  as well as to VOT when differentiating /t/ and /d/.

Finally, we used correlation analysis to examine the relation between gradiency of response and sensitivity to VOT and  $f_0$  on the VAS task. Sensitivity to the acoustic cue was represented by the subject-level VOT and  $f_0$  coefficients from the regression model separately constructed for each individual listener where VOT and  $f_0$  were independent variables (Morrison & Kondaurova, 2009). With respect to the VOT cue, listeners' coefficients were meaningfully correlated with their gradiency measure in both sessions (first session:  $r=0.88$ ,  $p<.0001$ ; second session:  $r=0.704$ ,  $p<.0001$ ). This positive correlation coefficient indicated that the more categorical listeners (larger quadratic slope) also tended to be more sensitive to VOT. The VOT coefficients between the two sessions were also significantly correlated ( $r=0.674$ ,  $p<.0001$ ), suggesting a systematic sensitivity to VOT within listeners.

However, the relationship between the gradiency measure and sensitivity to  $f_0$  was not systematic in the analysis of the VAS response patterns. The correlations between individuals'  $f_0$  coefficients and their gradiency were not statistically significant in the analyses of either session (first session:  $r=0.224$ ,  $p=.232$ ; second session:  $r=0.177$ ,  $p=.348$ ): The  $f_0$  coefficients across the two sessions were also not correlated ( $r=-0.117$ ,  $p=.536$ ). That is, the responses from the off-line VAS task failed to exhibit a systematic variability of  $f_0$  use in differentiating between voiced and voiceless stop consonants. We further investigated the relationship between gradiency in speech perception on the VAS task and listeners' sensitivity to the secondary cue of  $f_0$  by examining listeners' response on the AEM task, an on-line task.

### 3.2. AEM task

Fig. 4 shows the change in the estimated logit values of looking to /d/ over time as a function of different VOT and  $f_0$  conditions for the AEM task in the first session (see Appendix A and B for the figure and the model summary table based on the second session). This figure plots the estimated slopes from the mixed-effects models in six selected acoustic conditions. Table 1 presents the

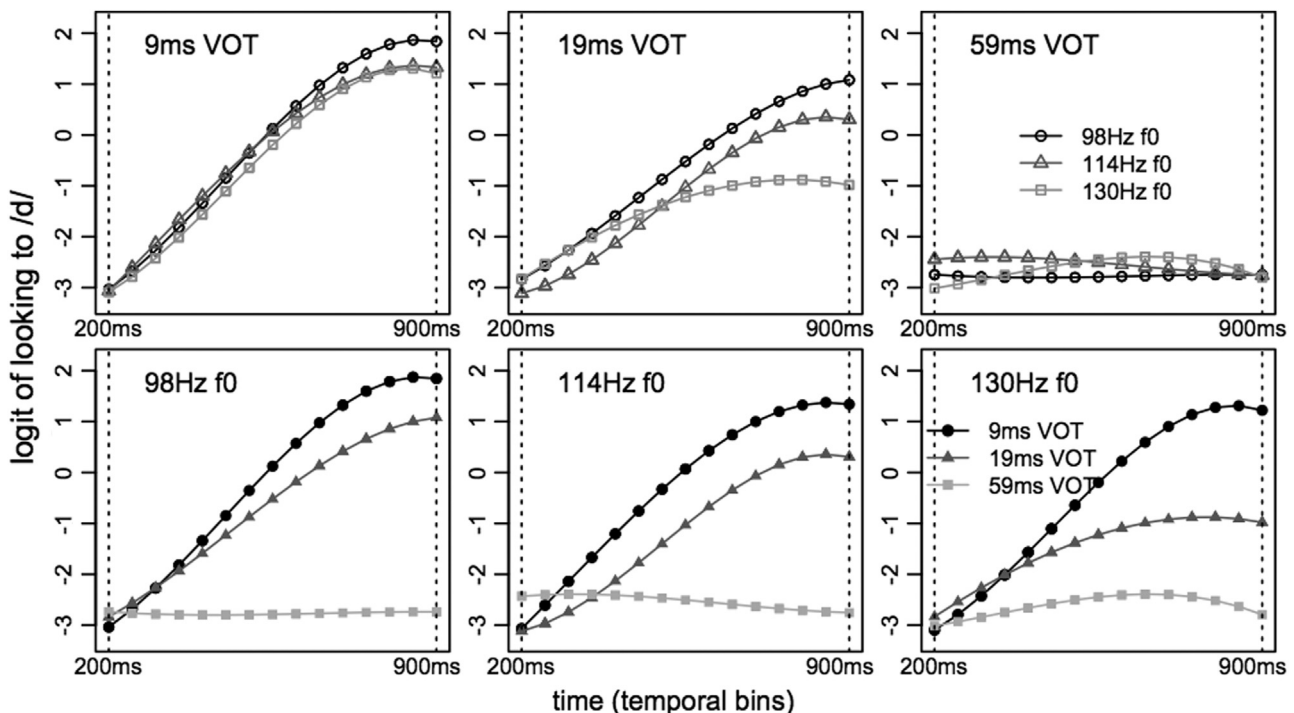


Fig. 4. Logit values of estimated linear slopes of looking to /d/ over time (temporal bins) in the AEM task (first session response only). The top row shows three VOT conditions in the three separate plots, with three line-types to indicate three different  $f_0$  conditions in each plot. The bottom row shows three  $f_0$  conditions in the three separate plots, with three line-types to indicate three different VOT conditions in each plot.

**Table 1**

Parameter estimations of the mixed effects time-series regression models in three different VOT conditions (9 ms, 19 ms and 59 ms). The lowest  $f_0$  value (98 Hz) was the reference condition of the model. Bold indicates coefficients with  $p$ -Values that were less than 0.05 and italic indicates coefficients with  $p$ -values that were less than 0.1.

	9 ms VOT		19 ms VOT		59 ms VOT	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept (98 Hz, reference $f_0$ )	<b>-0.857</b>	0.187	<b>-1.208</b>	0.182	-2.746	0.119
Linear slope (98 Hz)	<b>14.701</b>	1.125	<b>11.409</b>	1.273	-0.242	0.775
Quadratic slope (98 Hz)	-0.066	0.476	0.159	0.431	0.372	0.26
Cubic slope (98 Hz)	<b>-2.058</b>	0.305	<b>-1.164</b>	0.265	-0.158	0.167
Intercept (98 Hz: 114 Hz)	-0.212	0.16	<b>-0.361</b>	0.182	0.194	0.147
Intercept (98 Hz: 130 Hz)	-0.235	0.159	<b>-0.703</b>	0.181	0.024	0.148
Linear slope (98 Hz: 114 Hz)	-0.069	1.231	-1.933	1.395	-0.141	1.02
Linear slope (98 Hz: 130 Hz)	-1.635	1.227	<b>-4.197</b>	1.381	1.929	1.031
Quadratic slope (98 Hz: 114 Hz)	<b>-2.031</b>	0.673	<b>1.192</b>	0.566	<b>-1.136</b>	0.343
Quadratic slope (98 Hz: 130 Hz)	0.143	0.671	<b>-2.438</b>	0.539	<b>-1.302</b>	0.369
Cubic slope (98 Hz: 114 Hz)	<b>0.94</b>	0.424	<b>-0.985</b>	0.341	<b>0.473</b>	0.226
Cubic slope (98 Hz: 130 Hz)	-0.183	0.418	<b>0.928</b>	0.323	<b>-0.462</b>	0.234

**Table 2**

Parameter estimations of the mixed effects time-series regression models in three different  $f_0$  conditions (98 Hz, 114 Hz and 130 Hz). The shortest VOT condition (9 ms) was the reference condition of the model. Bold indicates coefficients with  $p$ -values that were less than 0.05 and italic indicates coefficients with  $p$ -Values that were less than 0.1.

	98 Hz $f_0$		114 Hz $f_0$		130 Hz $f_0$	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept (9 ms, reference VOT)	<b>-0.858</b>	0.158	<b>-1.066</b>	0.185	<b>-1.094</b>	0.153
Linear slope (9 ms)	<b>14.742</b>	1.139	<b>14.638</b>	1.113	<b>13.115</b>	0.989
Quadratic slope (9 ms)	-0.067	0.398	<b>-2.055</b>	0.418	0.071	0.376
Cubic slope (9 ms)	<b>-2.055</b>	0.255	<b>-1.14</b>	0.259	<b>-2.224</b>	0.227
Intercept (9 ms: 19 ms)	<b>-0.35</b>	0.187	<b>-0.503</b>	0.197	<b>-0.817</b>	0.175
Intercept (9 ms: 59 ms)	<b>-1.886</b>	0.189	<b>-1.475</b>	0.199	<b>-1.629</b>	0.179
Linear slope (9 ms: 19 ms)	<b>-3.327</b>	1.435	<b>-5.153</b>	1.383	<b>-5.898</b>	1.35
Linear slope (9 ms: 59 ms)	<b>-14.996</b>	1.462	<b>-15.108</b>	1.407	<b>-11.401</b>	1.407
Quadratic slope (9 ms: 19 ms)	0.229	0.556	<b>3.418</b>	0.542	<b>-2.35</b>	0.466
Quadratic slope (9 ms: 59 ms)	0.444	0.631	<b>1.339</b>	0.61	-1.028	0.602
Cubic slope (9 ms: 19 ms)	<b>0.885</b>	0.349	<b>-1.013</b>	0.329	<b>1.989</b>	0.277
Cubic slope (9 ms: 59 ms)	<b>1.902</b>	0.406	<b>1.446</b>	0.396	<b>1.626</b>	0.372

parameter estimations of the regression models associated with the three top panels of Fig. 4: the shortest (9 ms), mid-level (19 ms) and longest (59 ms) VOT conditions. Similarly, Table 2 shows the parameter estimations of the regression models associated with the three bottom panels of Fig. 4: the lowest (98 Hz), mid-level (114 Hz), and highest (130 Hz)  $f_0$  conditions. The patterns from the AEM task were mostly similar to the results for the VAS task in that they also showed that listeners were much more sensitive to changes in VOT than to changes in  $f_0$ . Across the different  $f_0$  conditions (bottom panels), stimuli with shorter VOT values consistently resulted in more looks to /d/ in each temporal bin than stimuli with longer VOT values. As summarized in Table 2, there were significant effects of the linear term between the 9 ms (reference category) and 59 ms VOT conditions across the three  $f_0$  models: The coefficients of the 59 ms VOT condition relative to the 9 ms VOT condition (Linear slope [9 ms:59 ms] in Table 2), were consistently negative and statistically significant at the level of  $p < 0.05$ , indicating that the curves were shallower for the 59 ms VOT condition. It can also be noted in Table 1 that the regression models for the 9 ms VOT and 59 ms VOT conditions did not show significant coefficients of the linear slope interaction between the reference  $f_0$  (98 Hz) and either of the other two  $f_0$  conditions (Linear slope [98 Hz:114 Hz] and [98 Hz:130 Hz] in Table 1), confirming that VOT values were the primary influence in determining looks to /d/ for the listeners.

While the slope differences of different  $f_0$  conditions were quite small in the shortest (9 ms) and the longest VOT (59 ms) conditions of Fig. 4 due to ceiling or floor effects, we identified one VOT condition in which listeners' sensitivity to  $f_0$  was relatively clear. In an ambiguous VOT condition (19 ms, shown in the top center panel), listeners looked more to /d/ over time as  $f_0$  decreased. That is, listeners attended to the  $f_0$  cue to voicing when the VOT cue was ambiguous. The slope for the highest  $f_0$  condition (130 Hz) in the 19 ms VOT condition model was shallower than that of the mid-level  $f_0$  (114 Hz) and that of the lowest  $f_0$  (98 Hz). Table 1 shows that in the 19 ms VOT condition, the linear slope coefficient of the highest  $f_0$  relative to the lowest  $f_0$  was significant ( $\beta = -4.19$ ,  $SE = 1.381$ ,  $p < 0.005$ ), indicating that listeners looked less to /d/ in the higher  $f_0$  condition relative to the lower  $f_0$  condition when the VOT was ambiguous.

We were interested in whether there was a relationship between the gradiency measure in the VAS task (slope coefficient from the quadratic term) and the measures of sensitivity to  $f_0$  in the AEM task. For this analysis, we examined individual listeners' response patterns in the AEM task in the conflicting vs. cooperating cue combinations of the highest and the lowest  $f_0$  conditions and in the ambiguous VOT conditions. Fig. 5 presents individuals' curves as well as the average fit from the model in the conflicting cue condition (130 Hz  $f_0$  and 9 ms VOT) and the cooperating cue condition (130 Hz  $f_0$  and 59 ms VOT) associated with the highest  $f_0$  in each of the two sessions. Fig. 6 shows the patterns in another conflicting cue condition (98 Hz  $f_0$  and 59 ms VOT) and cooperating

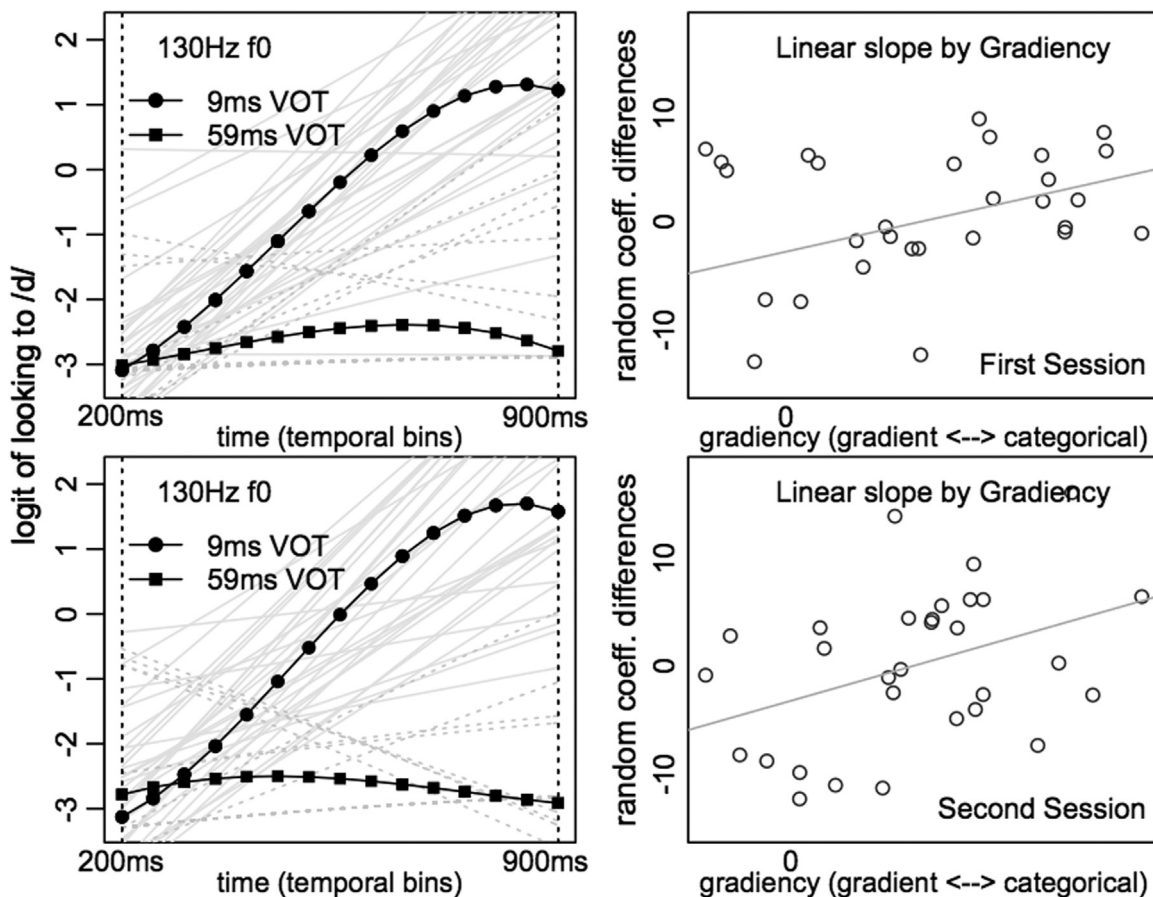


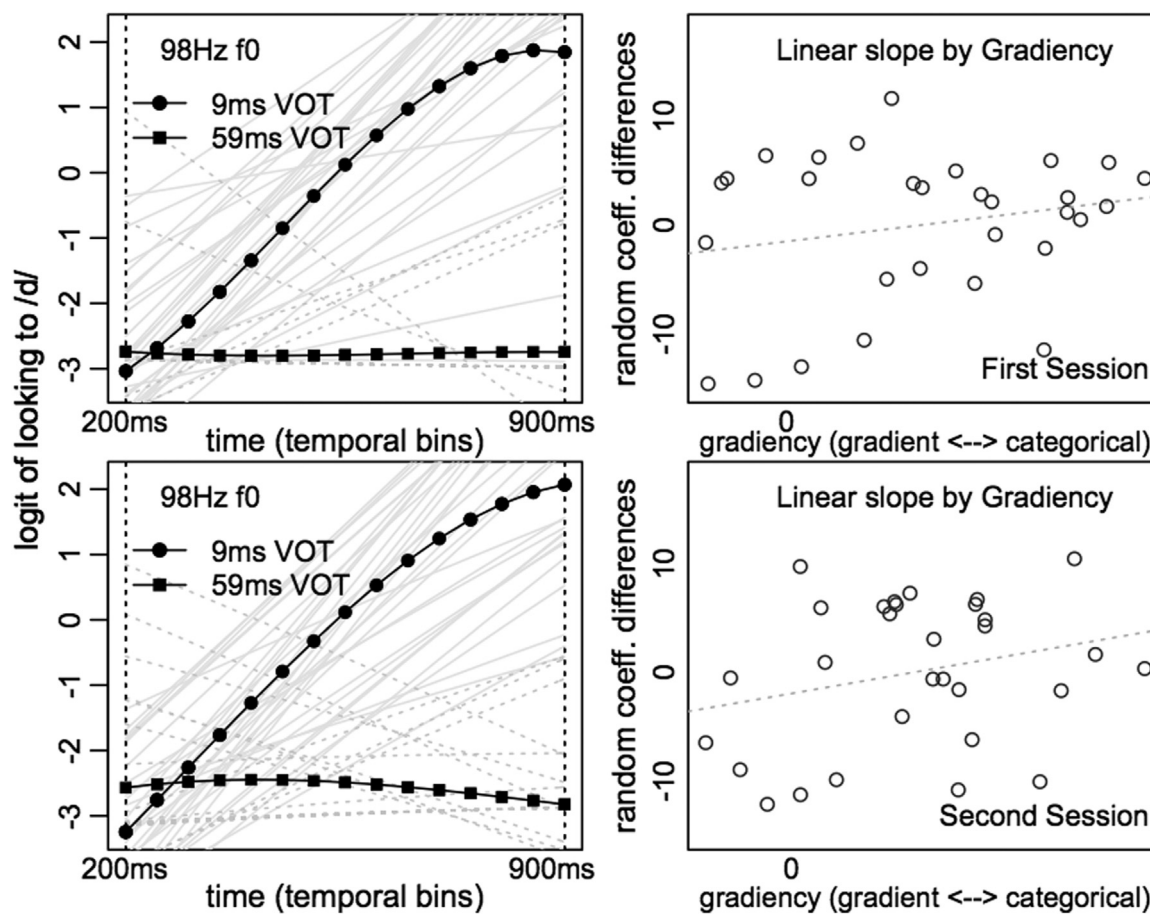
Fig. 5. The left panels show individual subject model fits for two of the time-series mixed effects regression models (9 ms and 59 ms VOT in the highest  $f_0$  condition [130 Hz]). Each curve indicates the model fit of an individual listener. The right panels show the difference between the slopes for these two curves plotted as a function of the gradiancy measure for each participant. Responses from the two test sessions are presented separately on the top (first session) and bottom (second session) rows.

cue condition (98 Hz  $f_0$  and 9 ms VOT) associated with the lowest  $f_0$  in each of the two sessions. Fig. 7 shows the ambiguous VOT conditions (19 ms and 28 ms VOTs combined) paired with the two  $f_0$  conditions (98 Hz and 130 Hz).

The left panels of Fig. 5 display the trajectories estimated from the regression models in the highest  $f_0$  condition. It can be observed that the slopes are generally steeper for the shorter VOT condition, but there is much individual variability in the conflicting cue condition. This individual variability in the trajectories was not completely random: a similar pattern appeared in the two AEM sessions, and individual listeners patterned consistently in these two sessions in that the linear slope coefficients of the 9 ms VOT condition were significantly correlated between the two sessions [ $r=0.416$ ,  $p=0.021$ ].

The right panels of Fig. 5 plot the linear slope coefficient differences between the 9 ms VOT and 59 ms VOT conditions calculated from the random effects at the subject-by-condition level (that is, the random effects coefficient of the 59 ms VOT condition was subtracted from that of the 9 ms VOT condition, see Section 2.4.2) as a function of the gradiancy measure from the VAS task (see Section 2.4.1). There was a significant positive correlation between the two dimensions (first session:  $r=0.369$ ,  $p=0.049$ ; second session:  $r=0.398$ ,  $p=0.029$ ). That is, the more gradient listeners tended to have smaller linear slope differences between the two VOT conditions while the more categorical listeners had larger linear slope differences between the two conditions. This pattern suggests that there was a greater influence of the  $f_0$  cue on voicing perception for gradient listeners. This result provides some evidence that individual differences in gradiancy of perception on the VAS task were related to individual differences in sensitivity to  $f_0$  on the AEM task.

Fig. 6 shows the trajectories estimated from the regression models in the lowest  $f_0$  condition (left panels). Similar to Fig. 5 of the highest  $f_0$  condition, the linear slopes were, on average, steeper for the 9 ms VOT condition than the 59 ms VOT condition. The variability observed in the individuals' slopes was not random in that the individuals' linear slope coefficients of the 9 ms VOT were correlated between the two AEM sessions [ $r=0.377$ ,  $p=0.039$ ]. Unlike the patterns in the highest  $f_0$  condition, however, the individual variability in the responses on the AEM task, represented by the linear slope coefficient differences between the 9 ms VOT and 59 ms VOT conditions, were not significantly correlated with the gradiancy measure from the VAS task (Fig. 6, right panels). While a visual inspection indicates an overall trend that more gradient listeners tended to have smaller slope differences between the two VOT conditions, the correlation coefficients were not statistically significant [first session:  $r=0.215$ ,  $p=0.251$ ; second session:  $r=0.258$ ,  $p=0.167$ ].



**Fig. 6.** The left panels show individual subject model fits for two of the time-series mixed effects regression models (9 ms and 59 ms VOT in the lowest  $f_0$  condition [98 Hz]). Each curve indicates the model fit of an individual listener. The right panels show the difference between the slopes for these two curves plotted as a function of the gradiency measure for each participant. Responses from the two test sessions are presented separately on the top (first session) and bottom (second session) rows.

To summarize, we found that there was a significant relation between gradiency and slope coefficient differences in the set of conflicting-cooperating cue conditions associated with the highest  $f_0$  but not in the other set associated with the lowest  $f_0$ . We speculate that the choice of “looks to /d/” as the dependent variable may be related to these inconsistent results. Given the dependent variable, the highest  $f_0$  with longer VOT cooperates for a percept of /t/ (where individual slopes are likely to be bounded at the floor), whereas the lowest  $f_0$  with shorter VOT cooperates for a percept of /d/. That is, it may be that this asymmetry might weaken the usefulness of the lowest  $f_0$  as an experimental condition examining how the conflicting  $f_0$  information affected individual listeners’ perception of the stops with a reference to their perception in the cooperating  $f_0$  information.

The results for the two stimulus conditions with ambiguous VOTs were similar to the results for the highest  $f_0$  condition. The left panels in Fig. 7 shows the trajectories based on the regression model estimations in the ambiguous VOT conditions (19 ms and 28 ms VOTs). While the linear slope differences between the higher  $f_0$  and lower  $f_0$  conditions are subtle based on the average model fits (thick lines), it can be observed that there was considerable variability in the subject-level curves. This individual variability was systematic and consistent within subjects across the multiple sessions; there was a significant correlation of the slope coefficients from the 98 Hz  $f_0$  condition between the two AEM sessions [ $r=0.374$ ,  $p=0.041$ ].<sup>2</sup> Furthermore, individual differences in the AEM task were correlated with individual differences in the gradiency measure from the VAS task. The right panels in Fig. 7 show the slope coefficient differences between the two  $f_0$  conditions (98 Hz and 130 Hz) estimated from the random effect plotted against the gradiency measure from the VAS tasks. Overall, there was a negative relationship between these two measures. Individuals who showed a more gradient response pattern on the VAS task had a greater linear slope differences between the two  $f_0$  conditions. Conversely, individuals who exhibited a more categorical response pattern on the VAS task had smaller linear slope differences between the two different  $f_0$  conditions. However, the correlation between the two measures was only significant for the responses from the second test session [first session:  $r=0.008$ ,  $p=0.96$ ; second session:  $r=-0.395$ ,  $p=0.030$ ]. Consistent with the pattern in

<sup>2</sup> It should be noted that the correlations between our working measures of  $f_0$  sensitivity (i.e., coefficient differences between shortest and longest VOTs at highest  $f_0$ ) and of  $f_0$  effect size (i.e., coefficient differences between lowest  $f_0$  and highest  $f_0$  at the ambiguous VOT) were not significantly correlated across the two test sessions. It is difficult to know how to interpret these findings, as researchers rarely examine consistency in eye gaze patterns across two test sessions and there are many differences between sessions, beginning with calibration. However, it is also important to note that although the difference scores were not correlated across the two test sessions, the linear slope coefficients themselves were. These slope coefficients came from the conditions where the slopes could potentially be steepest, thus allowing more room for individual variation.

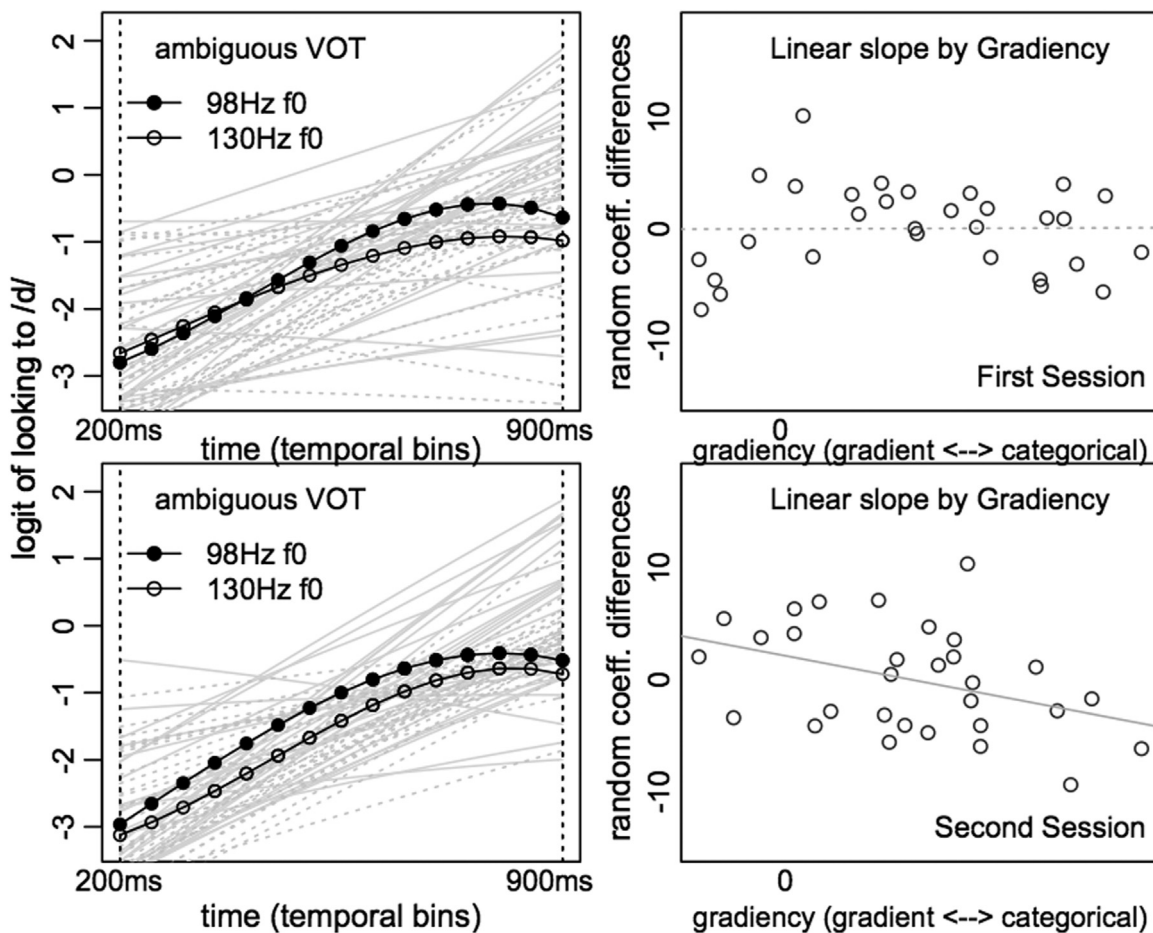


Fig. 7. The left panels show individual subject model fits for two of the time-series mixed effects regression models (98 Hz and 130 Hz  $f_0$ s in the ambiguous VOT conditions [19 ms and 28 ms VOTs]). Each curve indicates the model fit of an individual listener. The right panels show the difference between the slopes for these two curves plotted as a function of the gradiency measure for each participant. The two test sessions are presented separately in top (first session) and bottom (second session) rows.

the conflicting  $f_0$  cue condition illustrated in Fig. 5 (right panels), this result suggests that there was a greater effect of the  $f_0$  cue on voicing perception for the more gradient listeners as compared to the less gradient listeners. However, this result is quite tentative, given that it was significant only for the second test session.

### 3.3. Relationship between individual differences in executive function and speech perception patterns

We used partial correlations to examine whether any of the different measures of executive function were related to the measure of gradiency from the VAS task (slope coefficients of the quadratic regression models) or to the measure of sensitivity to  $f_0$  from the AEM task (linear slope differences between models for the 9 ms and 59 ms VOT conditions in the highest  $f_0$  condition and linear slope differences between models for the 98 Hz and 130 Hz  $f_0$  conditions in the ambiguous VOT condition). Tables 3 and 4 summarize the outputs of the partial correlation (two conditions: one, non-verbal IQ only partialled out and two, both non-verbal IQ and the other executive function measure partialled out).

The correlation tests with the gradiency measure were performed in two ways: (a) the gradiency measure was based on responses from all stimulus conditions, and (b) the gradiency measure was based on responses only from the ambiguous VOT conditions (19 ms and 28 ms VOTs). The output presented in Table 3(a) shows that, after applying the Bonferroni adjustment to the  $p$ -Value for multiple comparisons, the gradiency measure based on responses from all stimulus conditions was not consistently correlated with our measure of attention switching, response time on the Trail Making Task (TMT) for both of the two partial correlation tests: a significant correlation was only found between TMT and gradiency measured in the first session when IQ was partialled out. The partial correlation test with the other executive function test also yielded coefficients not reaching significance level ( $p < .0125$ ) in either session. Our measure of inhibition (response time on the Color-Word Naming task (CW)) was not correlated with the gradiency measure for either the first or the second test session for the partial correlation tests.

The results were similar when the gradiency measure was computed only over the responses from the ambiguous VOT conditions (Table 3b). TMT scores were not significantly correlated with the gradiency measure in either of the two partial correlation tests. Again, there was no significant correlation between our measure of inhibition and the gradiency measure. These results suggest that gradiency of speech perception and sensitivity to  $f_0$  were not related to higher level cognitive processing, at least for the measures assessed in this study.

**Table 3**

Output of the partial correlation tests between with each cognitive measure and the gradiency of response from the VAS task measured in two specific stimulus conditions: (a) all conditions and (b) ambiguous VOT conditions only. Results shown separately for data from the first and second sessions. Bold indicates  $p < .0125$ . [TMT: Trail Making Task, CW: Color-Word Naming test, IQ: non-verbal IQ test].

Correlation between gradiency &	Controlled covariates		First session		Second session	
			<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
(a) Gradiency measured in the all VOT conditions						
TMT	CW	IQ	−0.379	0.036	−0.289	0.123
TMT		IQ	<b>−0.477</b>	<b>0.004</b>	−0.344	0.056
CW	TMT	IQ	0.025	0.901	0.052	0.786
CW		IQ	−0.314	0.085	−0.201	0.283
(b) Gradiency measured in the ambiguous VOT conditions						
TMT	CW	IQ	−0.411	0.021	−0.423	0.017
TMT		IQ	−0.420	0.015	−0.363	0.042
CW	TMT	IQ	0.160	0.406	0.247	0.191
CW		IQ	−0.185	0.326	−0.084	0.660

**Table 4**

Output of the partial correlation tests between each cognitive measure and the *f0* sensitivity measure from the AEM task assessed in the two specific acoustic conditions: (a) highest *f0*, (b) lowest *f0* and (c) ambiguous VOT conditions. Results shown separately for data from the first and second sessions. [TMT: Trail Making Test, CW: Color-word naming test, IQ: non-verbal IQ test].

Correlation between <i>f0</i> sensitivity &	Controlled covariates		First session		Second session	
			<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
(a) Linear slope differences between 9 ms and 59 ms VOT in the highest <i>f0</i> condition						
TMT	CW	IQ	0.069	0.723	0.072	0.712
TMT		IQ	−0.165	0.384	0.025	0.893
CW	TMT	IQ	−0.271	0.150	−0.077	0.691
CW		IQ	−0.307	0.092	−0.038	0.842
(b) Linear slope differences between 9 ms and 59 ms VOT in the lowest <i>f0</i> condition						
TMT	CW	IQ	−0.145	0.453	0.14	0.446
TMT		IQ	0.034	0.856	0.301	0.100
CW	TMT	IQ	0.242	0.202	0.115	0.552
CW		IQ	0.199	0.290	0.287	0.118
(c) Linear slope differences between 98 Hz and 130 Hz <i>f0</i> in the ambiguous VOT condition						
TMT	CW	IQ	−0.085	0.660	−0.167	0.385
TMT		IQ	0.036	0.851	0.025	0.895
CW	TMT	IQ	0.160	0.406	0.263	0.162
CW		IQ	0.140	0.459	0.208	0.268

Similarly, the measure of sensitivity to *f0* from AEM task was not correlated with any of the executive function scores in any of the partial correlation tests. As presented in Table 4, there were no significant correlations between this slope measure and the executive function measures, regardless of whether sensitivity to *f0* was measured in the ambiguous VOT stimulus condition or in the highest or lowest *f0* condition for either of the two AEM sessions. At least for the measures assessed in this study, we did not find evidence that sensitivity to *f0* in the AEM task was related to cognitive processing.

#### 4. Discussion

The current study examined whether individual differences in how gradient listeners were in their perception of a stop voicing contrast was systematically related to differences in their sensitivity to a secondary acoustic cue. While the visual analogue scaling (VAS) task is designed to elicit attention to within-category differences, we found that there were large individual differences in performance on this task; response patterns for some listeners were much more gradient than response patterns for other listeners. A unique feature of this study was that we tried to quantify gradiency of performance on the VAS task. Using this measure, we were able to show that the gradiency of listeners' response patterns was consistent across two test sessions separated in time by about a week. This result suggests that gradiency of performance on the VAS task was not simply a task strategy that a listener adopts on a particular day, but was somehow associated with an individual listener's consistent pattern of speech perception. This finding is consistent with previous research that individual differences in speech are consistent within individual listeners (e.g., Idemaru & Holt, 2011; Schertz, Cho, Lotto, & Warner, 2015).

Evidence for a relation between individual differences in gradiency of performance on the VAS task and their use of a secondary acoustic cue was somewhat weaker. Listeners with a more categorical response pattern on the VAS task were more sensitive to VOT on this task than listeners with a more gradient response pattern. However, gradiency of response on the VAS task was not associated with greater sensitivity to *f0* on this task. Listeners with a more gradient response patterns on the VAS task tended to be

more sensitive to *f0* on an online eye-tracking task, although this result was not observed across all possible conditions. Kapnola, Winn, Kong, Edwards, and McMurray (2016) have found more robust support for the claim that more gradient listeners are more sensitive to the secondary cue on *f0* in a study that related gradency of response on a VAS task to use of *f0* on a 2AFC task. Such a finding is consistent with observations in previous research that there are individual differences in perceptual cue weighting patterns (Haggard et al., 1970; Stevens & Klatt, 1974; Walley & Carrell, 1983; Whalen et al., 1993; Idemaru et al., 2012; Toscano & McMurray, 2010). The current finding further supports the claim of Hazan and Rosen (1991) that individual differences in cue integration strategies are systematic. Hazan and Rosen (1991) found larger individual differences in reduced-cue conditions relative to full-cue conditions. Similarly, the current study found significant correlations between individual differences in gradency of perception and sensitivity to *f0* in conflicting *f0* cue conditions and in ambiguous VOT stimulus conditions. As noted above, these results are compatible with the proposal of Francis et al. (2008) for the *dynamic redistribution of attention*, although Francis and colleagues did not predict that there might be individual differences in whether attention would be dynamically redistributed. (See also Gordon et al. (1993) for the influence of a distractor task on attention to primary and secondary acoustic cues in speech perception.) Francis and colleagues hypothesized that attention to individual cues is reallocated dynamically by accommodating the usefulness of each cue to the specific listening condition. Under excellent listening conditions, listeners focus on primary acoustic cues to differentiate speech sounds. However, under more difficult listening conditions, such as listening to speech in noise or listening when attention is limited or when conflicting or ambiguous cues are present, listeners will also rely on secondary cues. The current study did not vary the demands of the overall listening conditions, but it was more difficult to differentiate between /t/ and /d/ in some of the stimulus conditions (e.g., in the competing cues condition or the ambiguous VOT conditions). We found that it was in these less ideal stimulus conditions that individual differences in attention to sub-phonemic acoustic details were observed.

Why was it the case that gradient listeners, rather than categorical listeners, were more flexible in utilizing redundant cues in difficult listening conditions, although this relationship was not robustly observed in all of the conditions we examined? It has been assumed in the literature with disordered populations such as children with dyslexia (e.g., Hazan & Barrett, 2000; Joanisse et al., 2000; Werker & Tees, 1987) that a categorical response pattern is preferable to a gradient response pattern and that individuals with a gradient response pattern have less robust phonological representations. By contrast, we found that individuals with a gradient response pattern on the VAS task were most able to attend to the secondary cue of *f0* on the AEM task in a challenging listening condition. With respect to this seemingly contradictory result, we point out that the VAS task – unlike the two-alternative forced choice (2AFC) paradigm of the classic categorical perception paradigm – is *designed* to elicit a gradient response pattern. We would argue that the “best” performers on the VAS task were those with a gradient response pattern and so it is not surprising that these listeners were also the most attentive to secondary cues in challenging listening conditions on the AEM task. Because speech cues are probabilistic and fine-grained subphonemic detail influences lexical processing (e.g., Clayards et al., 2008; McMurray et al., 2003), there are several advantages to a gradient response pattern. Everyday, non-laboratory listening situations are often quite noisy. There may be background noise or there may be interruptions to the signal (a door slam, a siren, etc.) and these results in perturbations to the speech signal. Fortunately, speech has redundant cues – if a durational cue such as VOT is obscured, another cue such as *f0* is still present. In these everyday listening situations, if individuals who have a more gradient response strategy are more likely to attend to multiple cues, then these listeners may have an advantage in word and sentence recognition. This suggests that a gradient response pattern may be preferable in real-life listening contexts.

In explaining individual differences in speech perception, we had hypothesized that gradency of performance on the VAS task would be related to general cognitive control. We tested this hypothesis by correlating our measure of gradency with performance on measures of inhibition and task shifting. We found little support for this claim and Kapnola et al. (2016) also did not find consistent relationships between gradency on a VAS task and measures of executive function. Furthermore, it should be noted that both the trail-making task and the color-word naming task are linguistic tasks of shifting and inhibition and our dependent variables are also linguistic as they both involve speech perception. It is possible that even the weak relationship between shifting and gradency that was present might not have been observed if we had used a non-linguistic measure of task shifting.

There are a number of reasons that we did not observe stronger or more consistent correlations between the executive function measures and the speech perception measures. It may be that our target population – normal hearing monolingual speakers – produced too narrow range of variability to statistically test the relationship. Also, it is possible that the individual differences observed here are related to other aspects of cognitive control, such as working memory, which was not examined in the current study. Another limitation of the current study was the relatively small number of stimuli (90) in each session. A larger number of stimuli might have resulted in a more robust measure of gradency for the VAS task and stronger correlations with the executive function measures. These are topics for future research.

Finally, the results of this study suggest some implications for L2 acquisition. We hypothesize that individuals with a more gradient response strategy in L1 speech perception will be better at learning a second language phonology relative to individuals with a more categorical response strategy in L1. It has been proposed that learning new cue weighting strategies is a crucial part of second language learning (Holt & Lotto, 2006; Iverson et al., 2003, 2005; Escudero, Bendersa & Lipski, 2009; Flege, Bohn, & Jang, 1997; Schertz et al., 2015, 2016; Silbert et al., 2015) as well as first language acquisition (Nittrouer, 1992, 2002; Nittrouer & Miller, 1997; Hazan & Barrett, 2000; Li, 2012; Walley & Carrell, 1983). Besides a group-level perceptual advantage or deficit of phonological structure between L1 and L2 (Best, 1995; Best & Tyler, 2007; Flege, 1995; Chang & Mishler, 2012), we suspect that individuals who attend more to secondary cues in order to supplement less informative primary cues in their first language are more attentive to how cue weighting differs in their second language and may be more successful L2 learners.

## 5. Conclusions

This study showed that some individuals were consistently more gradient than others in their differentiating between voiced and voiceless stop consonants on a VAS task. We also found that individuals with a more gradient response pattern also tended to be more attentive to  $f_0$ , a secondary cue to voicing. We speculate that a gradient response pattern in normal adults may be optimal in everyday listening contexts in a noisy environment.

## Acknowledgements

This work was supported by NIDCD grant R01 02932 and NSF grant BCS-0729140 to Jan Edwards and by NICHD P30 HD03352 grant to the Waisman Center. We thank all listeners who participated in this study. We are grateful to Mary E. Beckman, Elisabeth Bownik, Alia Dayne, Matt Goupell, Margarita Kaushanskaya, Benjamin Munson, Ryan Sovinski, and Jeewon Yoo for their help on many aspects of this study.

## Appendix A

See [Table A1](#).

**Table A1**

Parameter estimations of the mixed effects time-series regression models (1) in three different VOT conditions (9 ms, 19 ms and 59 ms) and (2) in three different  $f_0$  conditions (98 Hz, 114 Hz and 130 Hz) based on the response collected in the second session. The lowest  $f_0$  value (98 Hz) was the reference condition of the three VOT models and the shortest VOT condition (9 ms) was the reference condition of the three  $f_0$  models. Bold indicates coefficients with  $p$ -Values that were less than 0.05 and italic indicates coefficients with  $p$ -Values that were less than 0.1.

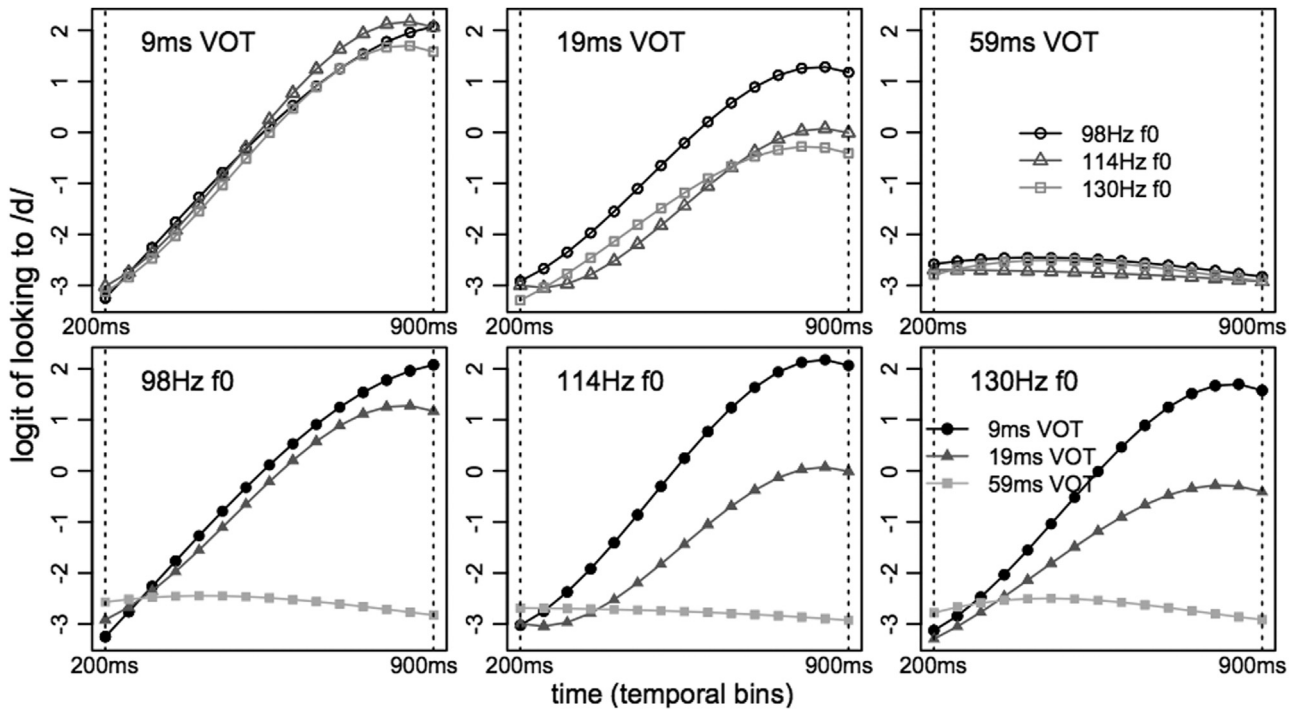
	9 ms VOT		19 ms VOT		59 ms VOT	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept (98 Hz, reference $f_0$ )	–1.025	0.178	–1.002	0.161	–2.608	0.114
Linear slope (98 Hz)	<b>16.657</b>	1.172	<b>12.085</b>	1.285	0.095	0.722
Quadratic slope (98 Hz)	–1.58	0.481	0.672	0.423	–1.024	0.269
Cubic slope (98 Hz)	–0.839	0.292	–2.458	0.254	0.163	0.181
Intercept (98 Hz: 114 Hz)	<b>0.354</b>	0.155	–0.623	0.226	–0.146	0.129
Intercept (98 Hz: 130 Hz)	0.099	0.155	–0.888	0.226	–0.153	0.129
Linear slope (98 Hz: 114 Hz)	–1.962	1.228	–5.394	1.465	–0.584	0.842
Linear slope (98 Hz: 130 Hz)	–2.739	1.228	–2.812	1.462	0.803	0.839
Quadratic slope (98 Hz: 114 Hz)	<b>2.809</b>	0.672	<b>3.091</b>	0.562	<b>0.841</b>	0.392
Quadratic slope (98 Hz: 130 Hz)	<b>2.311</b>	0.672	–1.208	0.549	–0.69	0.385
Cubic slope (98 Hz: 114 Hz)	–2.245	0.418	–0.642	0.33	–0.181	0.262
Cubic slope (98 Hz: 130 Hz)	–1.962	0.404	<b>0.915</b>	0.32	0.245	0.258

	98 Hz $f_0$		114 Hz $f_0$		130 Hz $f_0$	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept (9 ms, reference VOT)	–1.026	0.169	–0.674	0.152	–0.927	0.14
Linear slope (9 ms)	<b>16.655</b>	1.123	<b>14.719</b>	1.102	<b>13.917</b>	1.05
Quadratic slope (9 ms)	–1.562	0.43	<b>1.236</b>	0.387	<i>0.741</i>	0.391
Cubic slope (9 ms)	–0.852	0.261	–3.09	0.246	–2.809	0.233
Intercept (9 ms: 19 ms)	0.021	0.19	–0.951	0.177	–0.964	0.196
Intercept (9 ms: 59 ms)	–1.58	0.191	–2.078	0.18	–1.832	0.199
Linear slope (9 ms: 19 ms)	–4.573	1.402	–8.045	1.428	–4.644	1.421
Linear slope (9 ms: 59 ms)	–16.573	1.416	–15.236	1.466	–13.04	1.459
Quadratic slope (9 ms: 19 ms)	<b>2.236</b>	0.592	<b>2.536</b>	0.506	–1.274	0.501
Quadratic slope (9 ms: 59 ms)	0.519	0.639	–1.416	0.601	–2.464	0.597
Cubic slope (9 ms: 19 ms)	–1.616	0.357	–0.014	0.308	<b>1.265</b>	0.291
Cubic slope (9 ms: 59 ms)	<b>1.038</b>	0.411	<b>3.073</b>	0.393	<b>3.232</b>	0.381

## Appendix B

See [Fig. B1](#).



**Fig. B1.** Logit values of estimated linear slopes of looking to /d/ over time (temporal bins) in the AEM task (second session response only). The top row shows three VOT conditions in the three separate plots, with three line-types to indicate three different  $f_0$  conditions in each plot. The bottom row shows three  $f_0$  conditions in the three separate plots, with three line-types to indicate three different VOT conditions in each plot.

## References

- Abramson, A. S., & Lisker, L. (1985). Relative power of cues:  $f_0$  shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: essays in honor of Peter Ladefoged* (pp. 25–33). New York: Academic.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.
- Bates, D., Maechler, D., and Bolker, B. (2011). lme4: Linear mixed-effects models using Eigen and Eigen. R package version 0.999375-39. (<http://CRAN.R-project.org/package=lme4>) (date last viewed 08/30/13).
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Baltimore, MD: York Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: in honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins Publishing.
- Bialystok, E., & Craik, F. I. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, 19(1), 19–23.
- Boersma, Paul & Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.63, retrieved 24 January 2014 from (<http://www.praat.org>).
- Carney, A., Widen, C., & Viemeister, N. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 62, 961–970.
- Chang, C., & Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *Journal of the Acoustical Society of America*, 132, 2700–2710.
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–809.
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *The Delis-Kaplan executive function system* (pp. 1–218). San Antonio: The Psychological Corporation 1–218.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551–585.
- Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, 37(4), 452–465.
- Festman, J., Rodríguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, 6(1), 1.
- Flege, J., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 233–272). Baltimore, MD: York Press.
- Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 349.
- Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. J. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America*, 124, 1234–1251.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17(2), 172–179.
- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 152–162.
- Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25, 1–42.
- Gordon-Salant, S., & Fitzgibbon, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *Journal of the Acoustical Society of America*, 115, 1808–1817.
- Haggard, M. P., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading  $F_0$  cues in the voiced-voiceless distinction. *Journal of phonetics*.
- Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 47, 613–617.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28, 377–396.
- Hazan, V., & Rosen, S. (1991). Individual variability in response to cues to place contrasts in initial stops. *Perception & Psychophysics*, 49, 187–200.
- Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119(5), 3059–3071.
- Humes, L. E. (2002). Factors underlying the performance of elderly hearing aid wearers. *Journal of the Acoustical Society of America*, 112, 1112–1132.
- Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: the case of Japanese stops lengths. *Journal of the Acoustical Society of America*, 132, 3950–3964.

- Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1939–1956.
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r-l/ to Japanese adults. *Journal of the Acoustical Society of America*, 118, 3267–3278.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Ket-termann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, 47–57.
- 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(1), 30–60.
- Kapnoula, E. C., Winn, M. B., Kong, E., Edwards, J., & McMurray, B. (2016). Evaluating the sources and functions of gradience in phoneme categorization: an individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*. In preparation
- Karpicke, J., Conway, C.M., and Pisoni, D.B. (2007). Executive function, working memory, perceptual-motor skills, and speech perception in normal-hearing children: Some preliminary findings. Research on Spoken Language Processing Progress Report No. 28, Indiana University, 119–138.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test, Second edition* (pp. 1–136) Circle Pines, MN: American Guidance Service 1–136.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70, 419–494.
- Kim, Seongho (2012). ppcor: Partial and Semi-partial (Part) correlation. R package version 1.0. (<http://CRAN.R-project.org/package=ppcor>).
- Kong, E. J., and Edwards, J. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking (pp. 17–21). In *Proceedings of the XVIIth international congress of phonetic sciences*.
- Li, F. (2012). Language specific developmental differences in speech production: A cross-language acoustic study. *Child Development*, 83(4), 1303–1315.
- Lieberman, 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.
- Lieberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of components of certain speech and non-speech patterns. *Journal of Experimental Psychology: General*, 61, 379–388.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422.
- Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, 26, 708–715.
- Massaro, D. W., & Cohen, M. M. (1983). Integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 753–771.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics*, 53, 372–380.
- McMurray, B., Tanenhaus, M., Aslin, R., & Spivey, M. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface. *Journal of Psycholinguistic Research*, 32, 77–97.
- 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(6), 1609–1631.
- 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(1), 65–91.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *Journal of Memory and Language*, 59, 474–494.
- Morrison, G. S., & Kondaurova, M. V. (2009). Analysis of categorical response data: use logistic regression rather than endpoint-difference scores or discriminant analysis. *The Journal of the Acoustical Society of America*, 126(5), 2159–2162.
- Mullenix, J., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365–378.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 351–382.
- Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, 101, 2253–2266.
- Nittrouer, S. (2002). Learning to perceive speech: how fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, 112(2), 711–719.
- Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, 75, 224–230.
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285–290.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing URL(<http://www.R-project.org/>).
- Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 18(1), 12–14.
- Richardson, D., & Kirkham, N. (2004). Multimodal events and moving locations: eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, 133, 46–62.
- Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41, 71–80.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183–204.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*, 78(1), 355–367.
- Shukla, M., Wen, J., White, K. S., & Aslin, R. N. (2011). SMART-T: a system for fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods*, 43, 384–398.
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America*, 132, EL95–EL101.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford University Press.
- Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, 50, 99–119.
- Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 55, 653–659.
- Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. (1963). Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *The Journal of the Acoustical Society of America*, 35(11), 1900–1900.
- Toscano, J. C., & McMurray, B. (2010). Cue Integration with categories: weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34, 434–464.
- Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73, 1011–1022.
- Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2010). Colliding cues in word segmentation: the role of cue strength and general cognitive processes. *Language and Cognitive Processes*, 25, 402–422.
- Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology*, 41, 48–61.
- Whalen, D., Abramson, A., Lisker, L., & Mody, M. (1993). /f/ gives voicing information even with unambiguous voice onset time. *Journal of the Acoustical Society of America*, 93, 2152–2159.