The time course of perception of coarticulation

Patrice Speeter Beddor,[a] Kevin B. McGowan,[b] Julie E. Boland, Andries W. Coetzee, and Anthony Brasher
Department of Linguistics, University of Michigan, 611 Tappan Street, Ann Arbor, Michigan 48109

(Received 8 June 2012; revised 31 January 2013; accepted 8 February 2013)

The perception of coarticulated speech as it unfolds over time was investigated by monitoring eye movements of participants as they listened to words with oral vowels or with late or early onset of anticipatory vowel nasalization. When listeners heard [CṼNC] and had visual choices of images of CVNC (e.g., *send*) and CVC (*said*) words, they fixated more quickly and more often on the CVNC image when onset of nasalization began early in the vowel compared to when the coarticulatory information occurred later. Moreover, when a standard eye movement programming delay is factored in, fixations on the CVNC image began to occur before listeners heard the nasal consonant. Listeners' attention to coarticulatory cues for velum lowering was selective in two respects: (a) listeners assigned greater perceptual weight to coarticulatory information in phonetic contexts in which [Ṽ] but not N is an especially robust property, and (b) individual listeners differed in their perceptual weights. Overall, the time course of perception of velum lowering in American English indicates that the dynamics of perception parallel the dynamics of the gestural information encoded in the acoustic signal. In real-time processing, listeners closely track unfolding coarticulatory information in ways that speed lexical activation. © *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4794366]

PACS number(s): 43.71.An, 43.71.Es [CGC] Pages: 2350–2366

I. INTRODUCTION

Listeners are systematically exposed to what might be considered non-canonical forms of words. These forms include words with consonants or vowels that fail to reach their target trajectories, or words with expected consonants or vowels that are entirely absent from the acoustic signal. A major source of deviation from the presumed canonical form is the coarticulatory overlap of gestures for adjacent or nearby sounds, which gives rise to context-specific articulatory—and consequent acoustic—realizations of target segments. Coarticulatory overlap can mask or completely eliminate from the acoustic signal information potentially important to determining what a speaker has said (Stevens and Keyser, 2010). Coarticulation also enhances information available to listeners by providing cues to what is further up or down the speech stream.

Because coarticulation can both obscure and enhance perceptually useful acoustic information, it is perhaps unsurprising that theories of speech perception and of listener-speaker interactions differ in their accounts of the perceptual efficacy of coarticulated speech signals. Lindblom's hyper- and hypo-speech theory postulates that, in listener-directed speech, speakers will minimize coarticulation (by increasing articulatory effort) so as to maintain sufficient distinction between contrastive differences (Lindblom, 1990). Tatham and Moreton (2006) have similarly argued that speakers reduce coarticulation in order to increase the likelihood of being understood by listeners. Articulatory and acoustic investigations of this proposal have yielded mixed results. For example, clear speech studies (or studies comparing more- versus less-confusable words) demonstrate increased effort in clear speech (Moon and Lindblom, 1994; Matthies *et al.*, 2001; Bradlow, 2002), but they often show weak or no evidence of reduced coarticulation (Matthies *et al.*, 2001; Bradlow, 2002; Scarborough, 2004).

Several other theoretical approaches attend to the lawful, informative nature of the acoustic effects of overlapping articulations, which are arguably beneficial to listeners. Gestural theorists stipulate that listeners use coarticulatory cues in assigning a gestural parse to the acoustic input (e.g., Fowler, 1996). Many theorists—both gesturalists and non-gesturalists—more generally emphasize the communicative value of the perceptual coherence afforded by coarticulation (e.g., Whalen, 1984; Strange, 1989; Nearey, 1997; Hawkins, 2003). Models of spoken word recognition, such as TRACE, also incorporate listeners' use of anticipatory coarticulation to narrow in on the correct lexical item (Elman and McClelland, 1986).

We share the perspective of gesturalists and other theorists who assume that listeners, as active participants in processing the input acoustic signal, use the rich, time-varying information afforded by coarticulation to determine what speakers are saying. There is abundant evidence, some of which is reviewed below, that the acoustic consequences of coarticulatory overlap can be perceptually advantageous. In general, the findings show that, when provided with appropriate coarticulatory information, listeners respond more quickly and more accurately (e.g., Martin and Bunnell, 1981; Whalen, 1991; Connine and Darnieder, 2009) and can predict an upcoming or deleted segment with better than

[a] Author to whom correspondence should be addressed. Electronic mail: beddor@umich.edu
[b] Current address: Department of Linguistics, Rice University, Houston, TX 77005.

chance accuracy (Ostreicher and Sharf, 1976; Alfonso and Baer, 1982; Jenkins *et al.*, 1999). However, due in large part to methodological limitations until recent years, there is relatively little evidence that naturally produced, appropriate coarticulation facilitates perception during the initial processing of the utterance. If listeners use coarticulatory information in their moment-by-moment processing then, as the acoustic signal unfolds over time, listeners' perceptual assessments should evolve in ways that precisely use the time-varying information present in the signal. The present study investigates this prediction concerning real-time processing for coarticulatory vowel nasalization as perceived by native speakers of English.

Vowels in American English are typically produced with some degree of velum lowering when followed by a nasal consonant. This anticipatory lowering is especially extensive when the vowel and nasal are tautosyllabic (e.g., Krakow, 1999). Although anticipatory vowel nasalization exhibits idiolectal and dialectal variation, for many speakers nasalization is both temporally and spatially extensive (Solé, 1995) and in some cases extends throughout the entire vowel (e.g., Cohn, 1990; Beddor, 2009).

Unsurprisingly, English-speaking listeners use vowel nasalization in making judgments about vowel-nasal consonant sequences (VN) as opposed to vowel-oral consonant sequences (VC). Malécot (1960), in an early tape-editing experiment, deleted N from naturally produced CVNC(*er*) sequences in which C was voiceless (e.g., *camp*, *camper*) and found that listeners consistently identified stimuli as containing N on the basis of vowel nasalization alone. (Malécot's preliminary manipulations showed that, when C was voiced, N was required for a VNC percept.) Much more recently, Fowler and Brown (2000) showed that, although listeners are highly accurate in identifying consonants as oral or nasal regardless of whether the preceding vowel is oral or nasal, their reaction times were faster when the vowel had appropriate nasality ([CṼNə] and [CVCə]) than when, due to cross-splicing, vowel nasality was inappropriate ([CVNə] and [CṼCə]). MEG data indicate that, under passive auditory presentation of stimuli with appropriate and (cross-spliced) inappropriate nasality, participants' neural responses to oral C are delayed following a nasal vowel relative to following an oral vowel (Flagg *et al.*, 2006). Beddor's (2009) orthogonal variation of degree of vowel nasalization and duration of N revealed that, the temporally more extensive the coarticulatory information in the vowel is, the shorter the consonantal cue needed by listeners to identify stimuli as containing N (i.e., as *bent* and *bend* rather than *bet* and *bed*). Discrimination findings from that study point toward a similar perceptual trade-off between Ṽ and N.

Data from gating experiments provide further evidence that a coarticulatorily nasalized vowel informs English-speaking listeners about an upcoming nasal consonant. Warren and Marslen-Wilson (1987) cross-spliced the initial C(C)V and the coda N or C from minimal pair items such as *flown* and *float*, creating new inappropriate nasality sequences VN and ṼC, in addition to original appropriate ṼN and VC. As listeners were presented with increasingly long fragments of these stimuli, they identified the correct word more quickly for appropriate than for inappropriate nasality sequences. Comparable findings hold for gating experiments without cross-splicing: when presented with monosyllabic word fragments up through vowel offset, listeners were more likely to select a CVN word when the fragment had a nasal vowel than when it contained an oral vowel (Lahiri and Marslen-Wilson, 1991; Ohala and Ohala, 1995).

The clear pattern that emerges from these investigations is that listeners attend to anticipatory vowel nasalization in making decisions about a following consonant. The current investigation takes a further step and explores the extent to which the dynamics of perception are closely tied to the dynamics of the gestural information encoded in the acoustic signal. We investigated dynamic perception using a visual world paradigm (Allopenna *et al.*, 1998) in which participants' eye movements to a visual display were monitored as they listened to coarticulated speech. If listeners closely track the time-varying information in coarticulation in determining what speakers are saying, then their perceptual assessments, as shown by their visual fixations, will change as new acoustic information becomes available. An important advantage of tracking eye movements is that visual fixation information is continuously updated (with a time resolution of 5 ms), compared to the single judgments of a given stimulus or stimulus fragment in, for example, reaction time or gating studies. MEG data also provide excellent temporal resolution—recall that Flagg *et al.* (2006) found a delay in the neural response to an oral consonant when it was preceded by a nasal vowel. A crucial difference in the paradigms is that Flagg *et al.* studied neural responses to passive presentation of coarticulated speech, whereas looks to the correct item in the eye-tracking paradigm provide direct evidence of active use of coarticulatory information. This approach allows us to look for effects of coarticulatory nasalization during presentation of the vowel itself.

A small set of previous studies has monitored participants' eye movements in response to coarticulated speech. In this work, visual fixations in response to auditory stimuli containing coarticulatorily appropriate information are compared to fixations in response to cross-spliced stimuli with coarticulatorily inappropriate (mismatched) cues. For example, Dahan *et al.* (2001) investigated perception of V-to-C coarticulatory cues by cross-splicing the onset CV and coda C of CVC sequences (e.g., original *net*, *neck*, and *nep* yielded [nε_{k}t], [nε_{p}t], etc., where the subscript indicates the original coda consonant). They found that English-speaking participants were slowest and least accurate in fixating the image of the target item (*net*) when they heard a V with formant transitions for an upcoming C that would form a competitor word ([nε_{k}t]). Perceivers were more accurate when V contained transitional cues for a C that would form a non-word ([nε_{p}t]), and they were most accurate when V had coarticulatorily appropriate information ([nε_{t}t]). [See also Dahan and Tanenhaus (2004) for similar results for Dutch-speaking participants.] Extending the paradigm to C-to-C coarticulatory effects across a word boundary, Gow and McMurray (2007) found that participants fixated a target picture more quickly when the preceding word contained anticipatory place information for an upcoming target onset consonant (e.g., *green*$_b$ *boat*) than when the anticipatory cues were absent or misleading (e.g., *green*$_d$ *boat* spliced from *green dog* or *green*$_b$ *dog* spliced from *green boat*). Thus, results from existing studies show that inappropriate coarticulatory information slows participants' looks to a target image and temporarily increases incorrect looks to a competitor image. The current study extends this approach by focusing on participants' use of *appropriate* coarticulatory information. As explained below, in our study, relative speed of fixation is assessed not by comparing responses to matched versus (not spontaneously occurring) mismatched coarticulation, but rather via participants' responses to stimuli with differing degrees of (appropriate) coarticulation.

Two types of findings in the literature motivate our specific choice to study the perceptual time course of coarticulatory nasalization. First, American English-speaking listeners are exposed to variation in both the source and extent of vowel nasalization. Because different vowels have different intrinsic velum positions (Bell-Berti *et al.*, 1979) and velopharyngeal apertures (Moll, 1962; Clumeck, 1976), a somewhat open velopharyngeal port during vowel production in English is not an exclusively coarticulatory gesture. In addition, the extent of coarticulatory nasalization can be highly variable within and across speakers, and across prosodic and segmental contexts (Vaissière, 1988; Cohn, 1990; Bell-Berti, 1993; Krakow, 1993). In our own acoustic and aerodynamic analyses over the years, we have encountered (albeit rarely) American English speakers who have no measurable anticipatory nasalization, at least in certain pre-nasal contexts. Thus N is not the only source of vowel nasality, and coda N is not necessarily preceded by a robustly nasalized vowel. Such variation may have contributed to Lahiri and Marslen-Wilson's (1991) finding in their gating study that, while listeners were more likely to select a CVN word when the fragment had Ṽ than when it had V, the most frequent choice in both conditions was a CVC word. [But see Ohala and Ohala (1995) for a different pattern of results.] Although we fully expect listeners in our study to use coarticulatory nasalization in their perceptual decisions, the well-established variation for this coarticulatory process is a potentially informative source of context-specific and listener-specific differences.

A second, related motivation for investigating nasalization is to take advantage of the systematic effects of context on the temporal characteristics of ṼN sequences. In English, nasal consonants, like other sonorant consonants, are shorter when followed by voiceless than by voiced obstruents (Raphael *et al.*, 1975). Anticipatory nasalization of vowels preceding short (or, in some cases, absent) nasal consonants in pre-voiceless contexts tends to be temporally extensive (Malécot, 1960; Cohn, 1990). That is, the lowered velum gesture overlaps more with the vocalic configuration than with the consonantal constriction in VNC$_{\text{voiceless}}$ than in VNC$_{\text{voiced}}$ sequences in English (Beddor, 2009). As delineated below, we expect participants to use these precise, time-varying cues in real-time processing of English words.

The design of this study is an audio-visual task in which participants hear a CVC or CVNC auditory stimulus and see two pictures, a target image representing the auditory stimulus and a competitor image corresponding to a word that is minimally distinct from the target (e.g., target *bent*; competitor *bet* or *bend*). Saccades to the correct picture, launched during the vowel, provide direct evidence that coarticulatory knowledge is used to guide lexical access soon after the coarticulatory information becomes available. Our primary dependent measure is the proportion of fixations on the target picture at various time points as the word is being presented and shortly thereafter. We tested three main hypotheses.

Hypothesis 1 predicts that participants will use coarticulatory vowel nasalization to anticipate a nasal consonant, and they will use those cues shortly after that information becomes available in the unfolding acoustic signal. Stimuli with two temporal degrees of vowel nasalization were used to assess this hypothesis. When listeners hear a [CṼNC] stimulus, correct fixations on the target image should begin before the N is heard (taking into account the time it takes to program an eye movement; see below). Additionally, the mean latency of correct fixations should be shorter for stimuli with earlier onset of vowel nasalization.

Hypothesis 2 states that a coarticulatorily nasalized vowel will be a better indicator of an upcoming N than an oral vowel will be of an upcoming oral C. That is, we predict that presence of information about the velum lowering gesture will be more informative than its absence. This prediction is grounded in the perception literature, which suggests that listeners compensate for coarticulation, attributing coarticulatory effects to their source (e.g., Mann, 1980; among many others). Nasalization studies within this literature show that listeners report hearing the vowel in ṼN sequences as relatively oral (Kawasaki, 1986; Beddor and Krakow, 1999), an outcome that serves as evidence that perceiving a vowel as oral is not incompatible with an upcoming N. Data from gating studies with nasals also support this interpretation (Lahiri and Marslen-Wilson, 1991; Ohala and Ohala, 1995), as do the findings of Flagg *et al.* (2006), which showed a delay in the neural response to ṼC but not VN sequences, perhaps because, in English, a nasal vowel provides stronger predictive information than an oral one. [Conversely, the reaction time data of Fowler and Brown (2000) showed a greater delay in responses to VN than to ṼC sequences.]

Hypothesis 3 claims that listeners have detailed knowledge of the coarticulatory patterns of English and will use the voicing-dependent temporal patterns of ṼNC sequences in making lexical decisions. Participants are expected to use Ṽ to anticipate an upcoming N in both voiced and voiceless coda contexts. However, as the acoustic signal evolves over time, fixation patterns in these contexts may diverge. Because coarticulatory nasalization tends to be more extensive, and N shorter, before voiceless than before voiced codas, actual N should not be required for continued fixation of the CVNC$_{\text{voiceless}}$ image (e.g., *bent*). However, N may be required for continued fixation of the CVNC$_{\text{voiced}}$ image (*bend*), at least for some participants. This last hypothesis was tested in a "deleted N" condition, in which the auditory stimuli were [CṼC] items with voiceless or voiced coda consonants.

II. METHODS

A. Participants

Twenty-three participants were recruited from the University of Michigan community and either received credit in an undergraduate, introductory psychology class or were paid for participating in two testing sessions. All participants were adult, native English speakers with no known hearing deficits and normal (or corrected-to-normal) vision. An additional seven participants were recruited, but their results are not included in the data analysis due to difficulties calibrating the eye-tracker for particular individuals, failure to complete both testing sessions, or inattention to the task.

B. Stimuli

Target stimuli were five sets of minimal CV(N)C quadruplets whose members differed in presence of a nasal consonant and in final consonant voicing: *bet-bed-bent-bend*, *let-lead-lent-lend*, *set-said-scent-send*, *wet-wed-went-wend*, and *watt-wad-want-wand*. The original versions of the auditory stimuli were produced by an adult male native speaker of American English who is a trained phonetician and who has spent most of his life in Michigan. The speaker recorded multiple randomized repetitions of the 20 stimuli and 20 practice items embedded in the carrier phrase "Say __." From these repetitions, two instances of each word were selected for cross-splicing (see below); target words, with the carrier deleted, were selected on the basis of acoustic similarity (in f_0, duration, and vowel formant frequencies) to the other instance of that word and to other members of the quadruplet. Prior to any further manipulation, all recorded tokens were matched for peak intensity using a Praat (Boersma and Weenick, 2009) script.

To provide the necessary control over the time course of the coarticulatory information, the original versions were manipulated in Praat using waveform-editing techniques. All stimuli were cross-spliced. For each CVC-CVNC word pair matched for voicing (e.g., *bet*, *bent*), the initial CV portion was taken from an original CVC token (e.g., [b] and onset of [ɛ] in *bet* and *bent* were from the same *bet* token). The remainder of each CVC word was spliced from a second CVC token, whereas the remainder of each CVNC was taken from an original CVNC word (e.g., [ɛ̃nt] from *bent*). In all original CVNC words, vowel nasalization was clearly audible and acoustically identifiable. (Acoustic correlates of vowel nasalization included a decrease in amplitude of the waveform, and flattening and broadening of the low-frequency region in FFT spectra.)

For each CVNC stimulus, two temporal degrees of vowel nasalization were used to test the hypothesis that listeners attend to the coarticulatory cues shortly after they become available. For late onset of nasalization, the initial 60% of the vowel of the cross-spliced CVNC stimuli was from original CVC and the final 40% from original CVNC. For early nasalization onset, the proportion was 20% oral vowel and 80% nasal vowel. In most cases, the cross-splicing procedure entailed removing pitch pulses from the CV and VC or VNC portions of the original token to achieve the target vowel duration and degree of nasalization. In order to achieve 80% vowel nasalization in some of the early onset tokens, a small number of pitch pulses from the nasalized vowel were duplicated. Table I gives the resulting vowel and nasal consonant durations, averaged across the stimuli.

TABLE I. Average durations (in ms) of VN portions of target stimuli.

	Oral vowel	Nasal vowel	N
CVT	134.51		
CVD	185.96		
CVNT, Early onset $\tilde{V}$	26.85	100.85	51.33
CVNT, Late onset $\tilde{V}$	79.01	50.76	51.33
CVND, Early onset $\tilde{V}$	36.00	136.75	92.37
CVND, Late onset $\tilde{V}$	99.47	75.18	92.37

To test the hypothesis that listeners are sensitive to the context-specific coarticulatory patterns as $\text{CVNC}_{\text{voiceless}}$ and $\text{CVNC}_{\text{voiced}}$ sequences unfold, we further manipulated the heavily nasalized stimuli by excising the nasal consonant to create a deleted-N condition. The nasal consonant was identified from the waveform (by its relatively low amplitude and characteristic wave shape) and spectrographic displays, and excised at zero crossings. Thus, there were four types of auditory target stimuli: CVC, C$\tilde{V}$NC with late onset (40%) vowel nasalization, C$\tilde{V}$NC with early onset (80%) vowel nasalization, and C$\tilde{V}$C with early onset nasalization and N deleted. Excising and cross-splicing did not result in any audible signal discontinuities.

Each trial consisted of a single auditory stimulus and two visual stimuli. Table II specifies the pairs of target and competitor visual images and Table III lists the auditory stimulus conditions used with each type of visual pairing. (For the CVNT-CVND trials, auditory stimuli with deleted N were not included in order to keep the number of trials manageable.) The breakdown was 80 CVT-CVD (5 word pairs × 2 auditory stimuli × 8 repetitions), 80 CVNT-CVND (5 × 4 × 4), 100 CVT-CVNT (5 × 8 [CVT], 5 × 4 [C$\tilde{V}_{\text{early}}$NT], 5 × 4 [C$\tilde{V}_{\text{late}}$NT], 5 × 4 [C$\tilde{V}$T]), and 100 CVD-CVND (same as for CVT-CVNT), for a total of 360 test trials. Testing was conducted in two sessions, with half of the repetitions of each trial type occurring in each session. Each session also included 10 practice trials. The left versus right positions of the images in the visual display were counterbalanced across trials.

The critical visual stimuli were 20 black and white line drawings corresponding to each of the 20 critical words. The images were produced by a professional artist and were sized to fit within 5-in. square regions of a computer screen. Additional images were constructed for use in the practice trials.

TABLE II. Pairings of visual images.

CVT-CVD	CVT-CVNT	CVD-CVND	CVNT-CVND
bet-bed	bet-bent	bed-bend	bent-bend
let-lead	let-lent	lead-lend	lent-lend
set-said	set-scent	said-send	scent-send
wet-wed	wet-went	wed-wend	went-wend
watt-wad	watt-want	wad-wand	want-wand

TABLE III. Auditory stimuli for each type of visual pairing.

Visual pair	CVT-CVD	CVT-CVNT	CVD-CVND	CVNT-CVND
Auditory stimuli	[CVt]	[CVt]	[CVd]	[C$\tilde{\mathrm{V}}_{\mathrm{late}}$nt]
	[CVd]	[C$\tilde{\mathrm{V}}_{\mathrm{late}}$nt]	[C$\tilde{\mathrm{V}}_{\mathrm{late}}$nd]	[C$\tilde{\mathrm{V}}_{\mathrm{late}}$nd]
		[C$\tilde{\mathrm{V}}_{\mathrm{early}}$nt]	[C$\tilde{\mathrm{V}}_{\mathrm{early}}$nd]	[C$\tilde{\mathrm{V}}_{\mathrm{early}}$nt]
		[C$\tilde{\mathrm{V}}_{\mathrm{early}}$t]	[C$\tilde{\mathrm{V}}_{\mathrm{early}}$d]	[C$\tilde{\mathrm{V}}_{\mathrm{early}}$nd]

Because the design required CV(N)C quadruplets, stimuli could not be chosen on the basis of ease of identifying images. (In fact, when we informally asked some colleagues and students to identify the images, only *bed* was correctly identified by multiple respondents.) However, participants readily learned the labels for the visual stimuli in a familiarization task, described in Sec. II C. Lexical frequency of the critical words could also not be controlled in this design. Table IV gives the log frequencies based on the Corpus of Contemporary American English (COCA) (Davies, 2008). (The frequency of *lead*, which in these stimuli was the noun [lɛd], visually imaged by a pencil, is slightly inflated because the value also includes the noun [lid] plus sporadic verb hits.) Welch's *t*-tests, with log frequency as the dependent variable, show that there is not a significant relation between frequency and whether words are oral CVC or nasal CVNC ($p = 0.5089$), nor between frequency and whether words have voiced CV(N)D or voiceless CV(N)T codas ($p = 0.2309$). Moreover, as reported in Sec. III D, log frequencies do not predict listeners' fixation latencies.

C. Procedure

Participants were tested individually in two sessions. In the initial session, after collecting informed consent, participants learned the labels for each of the visual images used in the experiment. This was necessary because many of the target words were either difficult to represent as an image (e.g., *watt*) or were difficult to distinguish in the images (e.g., *bend* vs *bent*).

Labels for the images were learned in a two-part familiarization procedure that was repeated until perfect performance was achieved. Participants were first shown the images one at a time, randomly ordered, with the label written below each image. They were asked to memorize the labels by reading them aloud to the experimenter and explaining how each image might relate to each label; they viewed the images and labels at their own pace. Then participants were shown the images alone and required to produce the labels. Participants who failed to provide the correct label for one or more images repeated both the study phase and the test phase until they provided the correct label for all images. Most participants were 100% correct on the first test; no participant required more than two iterations of the familiarization procedure.

For the main part of the experiment, participants viewed pairs of the images on a computer screen and heard recorded instructions to look at one of the images. For example, participants saw images representing *bed* and *bend* and heard "Now look at bend." During this part of the experiment, we measured participants' eye movements to both images. Participants wore the headgear for an EyeLink II for eye movement monitoring and wore AKG k240 mkII headphones for presentation of the auditory stimuli. All visual stimuli appeared on the computer screen about 60 cm in front of the participant, on a height-adjustable table. The auditory and visual stimuli were presented using SR Research Experiment Builder software.

Participants were seated in a chair while the experimenter positioned the Eyelink headgear and adjusted it for optimal tracking and a secure fit. The Eyelink has binocular eye cameras mounted on a headband, with a sampling rate of 500 Hz. When necessary, table height was adjusted so that the participant's eye level was in the middle of the top half of the computer screen. Room lighting was low and indirect. Before the experiment began, the experimenter performed a calibration procedure, which was repeated if necessary until criterion was reached for both eyes. Data from the best eye were stored and used for analysis. Before each trial, a drift correction procedure was performed. A 5-min break was enforced at the halfway point of each session after which the calibration process was repeated prior to continuing.

Each trial lasted about 10 s and consisted of the following sequence of events. First, the two images appeared in the left and right halves of the screen. Each image was fit inside a 5-in. 72 dpi square; the screen was 1024 × 768 pixels. Participants heard the instruction "Look at the pictures." After 2 s, a fixation cross appeared in the center of the screen, as shown in the sample screen shot for *set* and *scent* in Fig. 1. Participants heard "Fixate cross. (pause) Now look at." At this point, participants heard the critical auditory stimulus and the fixation cross disappeared. The trial ended 2 s later. Participants completed ten practice trials before starting the experiment proper.

The second testing session was necessary due to the large number of trials. Half of the data collection occurred in

TABLE IV. Log (base 2) frequencies of target words based on COCA (Davies, 2008).

CVT		CVD		CVNT		CVND	
Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
bet	13.99	bed	15.88	bent	13.49	bend	13.08
let	17.96	lead	13.59	lent	8.81	lend	11.94
set	17.28	said	20.08	scent	12.63	send	15.25
watt	9.85	wad	10.12	want	18.52	wand	9.98
wet	14.13	wed	6.74	went	17.50	wend	6.87

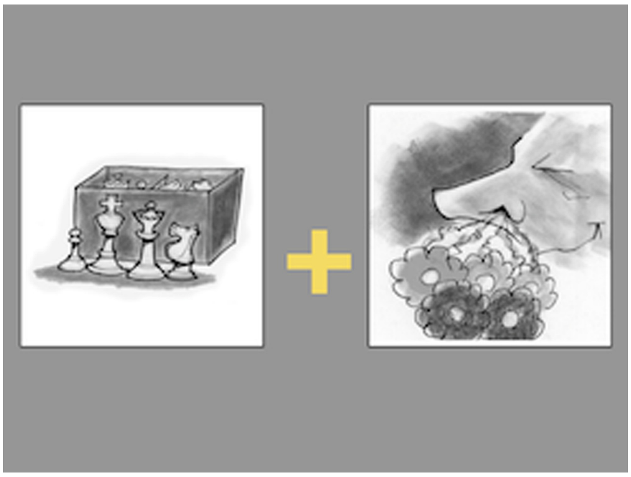

FIG. 1. (Color online) Sample screen shot for the trial *set–scent*.

each testing session, which were typically a few days apart. On the second day of testing, participants were again familiarized with the labels for the images and tested to ensure 100% accuracy on the image-labeling task. Then another eye tracking session was conducted.

D. Measures and predictions

Participants' eye movements to the two images on the computer screen were monitored during a critical interval starting from the onset of the target stimulus item (e.g., the onset of [b] in *bend*) and lasting 1000 ms. The measures were latency of initial correct fixations and the proportion correct fixations over time. Only trials on which there was a fixation within 1000 ms following the onset of the target word were included in the data analysis. Latency was measured from the onset of the critical word until the eye gaze first entered the 5-in. square screen region containing the image named by the critical word. Proportions were computed for 20-ms temporal bins, beginning at the onset of the critical word. A proportion of 0 during the bin starting at 200 ms means that there were no trials in that condition with a fixation on the target during any portion of the 20-ms interval from 200 to 220 ms. A fixation was counted as a target (correct) fixation if it fell within the 5-in. square region containing the target image. A proportion of 0.50 means that 50% of the trials in the condition included a target fixation that ended, began, or continued throughout that 20-ms interval.

The hypotheses stated in Sec. I were operationalized through four main types of comparisons across trial types; these are summarized in Table V. Hypothesis 1 states that listeners will use coarticulatory vowel nasalization to anticipate an upcoming nasal consonant and they will do so soon after the information becomes available. When participants hear [CṼNC], latency to fixate the corresponding picture should be shorter when the competitor visual image represents a word that lacks a nasal consonant than when the competitor image is of a word with a nasal consonant. That is, for auditory [CṼNC] (e.g., *bent*), latency should be shorter for visual CVNC-CVC (*bent-bet*) trials than for CVNT-CVND (*bent-bend*) trials. We speculated that latencies are shorter for CVNC-CVC trials at least in part because vowel nasalization serves as a disambiguating cue; this is not true when the visual competitor is also CVNC. However, even without attending to vowel nasalization, listeners should fixate CVNC more quickly when the competitor is CVC than when it is another CVNC image because hearing the nasal consonant would differentiate images of CVNC and CVC words but not images of CVNT and CVND words. As a result, it is necessary to establish that looks to target CVNC items occur during the vowel portion of the stimulus. Comparisons between early and late onset of nasalization are critical to Hypothesis 1. If listeners' initial looks to CVNC words are based on vowel nasalization in [CṼNC], the earlier the onset of the coarticulatory cue, the faster and more accurately participants should respond. Thus, for visual CVNC-CVC (*bent-bet*) trials, latency of initial correct fixations should be shorter and looks to the CVNC item should be more frequent when the auditory stimulus is [C$\tilde{\text{V}}_{\text{early}}$NC] than when it is [C$\tilde{\text{V}}_{\text{late}}$NC]. Moreover, if listeners use the coarticulatory information nearly as soon as it occurs, looks to CVNC items should begin shortly after 200 ms after the onset of vowel nasalization. This estimate assumes that it takes roughly 200 ms to program and launch an eye movement (Dahan *et al.*, 2001).

TABLE V. Predicted fixation patterns for planned comparisons of stimulus pairs, illustrated for the *bet-bed-bent-bend* set. **Bold** indicates the auditory stimulus; x–y indicates the visual pairing. "<" and "=" refer to relative latencies of initial correct fixations.

	Predicted fixation patterns		
Baseline comparisons	$\textbf{be}_{\textbf{early}}\textbf{nt}$ – bet	<	$\textbf{be}_{\textbf{early}}\textbf{nt}$ – bend
	$\textbf{be}_{\textbf{late}}\textbf{nt}$ – bet	<	$\textbf{be}_{\textbf{late}}\textbf{nt}$ – bend
	$\textbf{be}_{\textbf{early}}\textbf{nd}$ – bed	<	$\textbf{be}_{\textbf{early}}\textbf{nd}$ – bent
	$\textbf{be}_{\textbf{late}}\textbf{nd}$ – bed	<	$\textbf{be}_{\textbf{late}}\textbf{nd}$ – bent
Hypothesis 1	$\textbf{be}_{\textbf{early}}\textbf{nt}$ – bet	<	$\textbf{be}_{\textbf{late}}\textbf{nt}$ – bet
	$\textbf{be}_{\textbf{early}}\textbf{nd}$ – bed	<	$\textbf{be}_{\textbf{late}}\textbf{nd}$ – bed
Hypothesis 2	**bet** – bent	=	**bet** – bed
	bed – bend	=	**bed** – bet
Hypothesis 3	**be(n)t** – bet more correct fixations than **be(n)d** – bed		

Hypothesis 2 predicts that an oral vowel, unlike a nasal vowel, will not serve as a strongly disambiguating cue. If this hypothesis is upheld, then, when participants hear a [CVC] stimulus (*bet*), the latency for trials in which the visual options are CVC-CVNC (*bet-bent*) will not differ significantly from those in which the visual options are CVT-CVD (*bet-bed*).

Trials with auditory [CṼC], with the nasal consonant excised, test Hypothesis 3, which predicts that listeners' use of coarticulatory nasalization will be sensitive to context-dependent timing patterns in English. The pattern of interest here is that, in production, the lowered velum gesture overlaps more with the vowel and less with the consonantal constriction in voiceless than in voiced contexts. When participants hear [CṼC] and see CVNC-CVC visual images, the latency of initial fixations on the CVNC image should be similar for the two voicing contexts but, across the time course of the target word, the CVNC image should elicit overall more fixations when the coda consonant is voiceless ([bɛ̃t]) than when it is voiced ([bɛ̃d]).

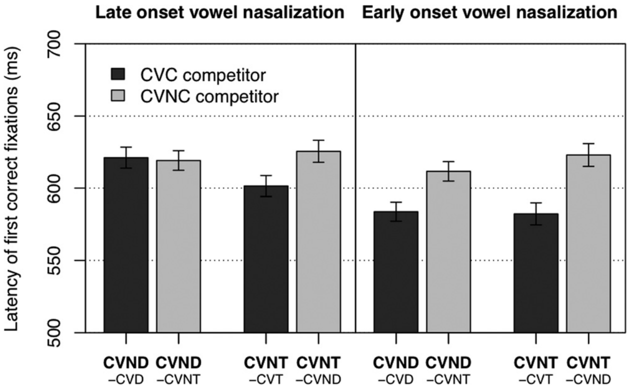

FIG. 2. Mean latency of first correct fixations on trials with auditory [CṼNC] according to vowel nasalization (left, right panels), competitor picture (bar type), and coda voicing (voiced: left set of bars in each panel; voiceless: right set). Errors bars represent standard error of the mean.

III. RESULTS

The predictions summarized in Table V are largely, but not entirely, upheld by the fixation patterns for the different trial types. Results are presented according to type of auditory stimulus, [CṼNC], [CVC], and [CṼC]. Throughout the presentation of the results, the visual target corresponding to the auditory stimulus is listed first (e.g., in a CVND-CVD trial, the auditory stimulus was [CṼND]).

A. Auditory [CṼNC] trials

Figure 2 gives the mean latencies of the first correct fixations on the trials in which participants heard a nasalized vowel followed by a nasal consonant ([CṼNC]). Overall, as expected, latencies were shorter when the competitor image was of a CVC item (dark bars) than when it was of another CVNC item that differed in coda voicing (light bars). Latencies were also overall shorter when participants heard [CṼ$_{\text{early}}$NC] (Fig. 2, right) than when they heard [CṼ$_{\text{late}}$NC] (Fig. 2, left), but only for visual CVND-CVD and CVNT-CVT trials. As predicted, latencies for early versus late nasalization are nearly identical for the CVND-CVNT and CVNT-CVND trials. Thus, participants fixated the correct image more quickly only when coarticulatory nasalization was a disambiguating cue. Additionally, although more extensive coarticulation was expected to trigger greater facilitation, both early and late nasalization should result in some facilitation. That is, even in the late nasalization condition, CVNC-CVC trials should have shorter latencies than CVNC-CVNC trials. This prediction was upheld for [CṼ$_{\text{late}}$NT] but not for [CṼ$_{\text{late}}$ND] stimuli. Unexpectedly, when listeners heard [CṼ$_{\text{late}}$ND], fixation latencies were the same for CVND-CVD and CVND-CVNT visual pairings.

A linear mixed model was computed on first correct fixation latencies for auditory [CṼNC] stimuli. The model was fit using the lmer() function in the lme4 package in R (Bates *et al.*, 2011). Fixed effects were Degree of Nasalization (early, late) and Visual Competitor (CVD, CVT, CVNT, CVND); participant and item were included in the model as random intercepts. The linear mixed model does not supply an omnibus test but rather directly models the main effects as paired comparisons using *t*-tests. Although interaction effects are included in the model, they are not presented here because the paired comparisons are sufficiently informative. *p* values were estimated using Markov Chain Monte Carlo simulations using the pvals.fnc() function from the languageR package (Baayen, 2008).

Table VI gives the model results for the comparisons that tested the baseline comparisons and Hypothesis 1. Comparisons conducted separately for the [CṼ$_{\text{early}}$NC] and [CṼ$_{\text{late}}$NC] auditory prompts tested the prediction that listeners use coarticulatory vowel nasalization to anticipate an upcoming nasal consonant. When participants heard [CṼ$_{\text{early}}$NC], fixation latencies were significantly shorter when the visual competitor was an image of a CVC word (compared to a CVNC word; see paired dark and light bars in Fig. 2). This outcome held for both the voiced and voiceless conditions. However, when participants heard [CṼ$_{\text{late}}$NC], fixations were significantly shorter when the visual competitor was a CVC item only for the voiceless condition. The additional prediction that the earlier the onset of coarticulatory nasalization, the faster participants will respond is upheld for both voicing conditions. That is, for both [CṼND] and [CṼNT] auditory prompts, earlier vowel nasalization led to shorter latencies when nasalization was potentially disambiguating (visual CVNC-CVC trials; compare corresponding dark bars in left and right panels of Fig. 2). As expected, earlier vowel nasalization did not influence response latencies when the coarticulatory information did not help differentiate the visual options (visual CVND-CVNT trials; corresponding light bars on left and right).

The effect of degree of nasalization emerges not only for initial correct fixations, but continues to hold as the

TABLE VI. Paired comparisons testing Hypothesis 1 from linear mixed model fit to first correct fixation latencies for auditory [CṼNC]. **Bold** indicates the auditory stimulus in each comparison.

	Comparison	Estimate (β)	t	p
NC vs C competitor, Early Nasalization	**CVND$_{\mathbf{early}}$**-CVNT: **CVND$_{\mathbf{early}}$**-CVD	30.52	3.64	0.0001
	CVNT$_{\mathbf{early}}$-CVND: **CVNT$_{\mathbf{early}}$**-CVT	39.95	4.64	0.0001
NC vs C competitor, Late Nasalization	**CVND$_{\mathbf{late}}$**-CVNT: **CVND$_{\mathbf{late}}$**-CVD	0.32	0.04	0.9618
	CVNT$_{\mathbf{late}}$-CVND: **CVNT$_{\mathbf{late}}$**-CVT	25.66	2.97	0.0040
Early vs Late, Voiced Coda	**CVND$_{\mathbf{early}}$**-CVD: **CVND$_{\mathbf{late}}$**-CVD	39.31	4.67	0.0001
	CVND$_{\mathbf{early}}$-CVNT: **CVND$_{\mathbf{late}}$**-CVNT	8.47	1.00	0.3178
Early vs Late, Voiceless Coda	**CVNT$_{\mathbf{early}}$**-CVT: **CVNT$_{\mathbf{late}}$**-CVT	21.41	2.51	0.0122
	CVNT$_{\mathbf{early}}$-CVND: **CVNT$_{\mathbf{late}}$**-CVND	7.13	0.81	0.4252

acoustic signal unfolds over time. Figure 3 provides, for [C$\tilde{V}_{\text{early}}$NC] and [C$\tilde{V}_{\text{late}}$NC] auditory prompts, the mean proportion of correct fixations, over time, of the CVNC image for the voiced and voiceless conditions. The proportions could conceivably rise and fall over time, as participants fixate the target and then look away, yet once participants had looked at the target they tended to remain looking at that image for the remainder of the 1-s interval during which eye movements were tracked.

Although roughly 90% of participants' fixations converge on the CVNC image within 1 s regardless of whether onset of anticipatory nasalization is early or late, for trials in which the competitor image is of a CVC word (filled versus open circles) there appear to be comparatively more correct fixations when nasalization began early than when it began late at many of the time points represented in Fig. 3. The earlier convergence of the early and late nasalization functions in the CVNT than in the CVND condition is not surprising, given shorter V and N durations in the voiceless condition.

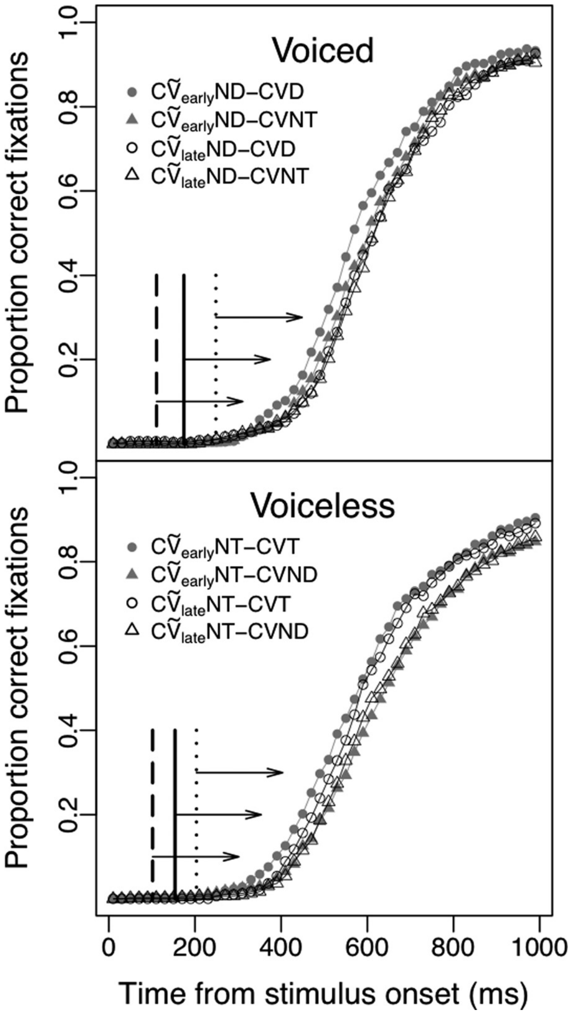

FIG. 3. Pooled proportion correct fixations on trials with auditory [C$\tilde{V}$NC] according to degree of vowel nasalization (filled vs unfilled symbols) and competitor image (circles vs triangles) for voiced and voiceless conditions. Short dashed line: early vowel nasalization onset; solid line: late vowel nasalization onset; dotted line: N onset. Arrows indicate 200 ms eye movement programming delay.

A linear mixed model with a logit link function was fit to the proportion correct fixations for [C$\tilde{V}$NC] stimuli. The model was calculated over a 360 ms window whose onset was 200 ms after the location of the splice for early vowel nasalization onset. For [C$\tilde{V}_{\text{late}}$NC] stimuli, window onset was also 200 ms after the point at which early onset nasalization would have begun; e.g., onsets for *send*$_{\text{late}}$ and *send*$_{\text{early}}$ were the same, so that onset corresponded to the earliest splice for each word type, as in Fig. 4. Offset of the window was chosen to match the average duration of the stimuli (plus 200 ms) following the early vowel nasalization splice. Table VII gives the model results for the comparisons of interest. Within the target time window, the proportion correct fixations of the CVNC image was significantly higher with early than with late vowel nasalization for both voicing conditions when the competitor image was CVC. It was predicted that the early-late nasalization comparison would not be significant when the competitor image was another CVNC word, and indeed it was not in the voiceless context ($p = 0.6290$). The early-late comparison unexpectedly reached significance for the voiced CVND trials when the competitor image was of a CVNT word, but effect size was small (mean difference of 2% compared to a mean difference of over 7% for the CVND-CVD early-late comparison).

The effect of visual competitor also generally holds for proportion correct fixations, with an overall higher proportion of target fixations when the vowels of the two images would be expected to differ in nasality (i.e., CVNC-CVC trials; circles in Fig. 3) than when they should have the same nasality (CVNC-CVNC trials; triangles). The paired comparisons reported in Table VII show that this effect holds for

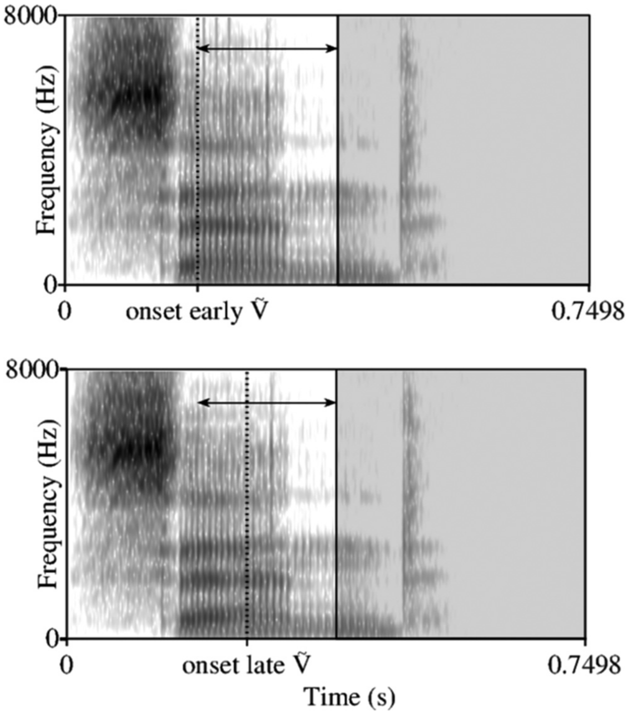

FIG. 4. Illustration, for *send*, of 360 ms window (shaded portion) used for the linear mixed model fit to proportion correct fixations for [C$\tilde{V}$NC] stimuli. Dotted line marks onset of vowel nasalization; arrow marks 200 ms delay from onset of early vowel nasalization.

TABLE VII. Paired comparisons testing Hypothesis 1 from linear mixed model fit to proportion correct fixations for auditory [C$\tilde{\text{V}}$NC]. **Bold** indicates the auditory stimulus in each comparison.

	Comparison	Estimate (β)	z	p
NC vs C competitor, Early Nasalization	**CVND$_{\text{early}}$**-CVNT: **CVND$_{\text{early}}$**-CVD	0.283	7.98	0.0001
	CVNT$_{\text{early}}$-CVND: **CVNT$_{\text{early}}$**-CVT	0.539	14.71	0.0001
NC vs C competitor, Late Nasalization	**CVND$_{\text{late}}$**-CVNT: **CVND$_{\text{late}}$**-CVD	0.008	0.23	0.8221
	CVNT$_{\text{late}}$-CVND: **CVNT$_{\text{late}}$**-CVT	0.297	7.91	0.0001
Early vs Late, Voiced Coda	**CVND$_{\text{early}}$**-CVD: **CVND$_{\text{late}}$**-CVD	0.426	11.79	0.0001
	CVND$_{\text{early}}$-CVNT: **CVND$_{\text{late}}$**-CVNT	0.134	3.62	0.0003
Early vs Late, Voiceless Coda	**CVNT$_{\text{early}}$**-CVT: **CVNT$_{\text{late}}$**-CVT	0.261	7.37	0.0001
	CVNT$_{\text{early}}$-CVND: **CVNT$_{\text{late}}$**-CVND	0.019	0.48	0.6290

three of the four conditions. The one exception is that, as was the case for initial correct fixation, there is not a significant effect of visual competitor in the voiced condition when onset of coarticulatory nasalization is late.

To more precisely measure the time course of perception of coarticulatory nasalization, we determined the time (averaged over 20 ms time bins) at which the proportion correct fixations on CVNC-CVC trials first differed significantly from fixations on CVNC-CVNC trials for early onset of vowel nasalization. To accomplish this, we iterated over the time bins, fitting a generalized linear mixed model to each 20 ms interval. If significant divergence occurs before N onset, this indicates that, in the CVNC-CVC condition, listeners began targeting looks at the CVNC pictures based on coarticulatory nasalization. For example, for [C$\tilde{\text{V}}_{\text{early}}$NT] stimuli, vowel nasalization began, on average, 102 ms after stimulus onset and N began an average of 203 ms after stimulus onset. Comparison of the relevant response curves in Fig. 3 (filled circles versus triangles, lower panel) shows that the curves begin to diverge 340 ms from stimulus onset; the difference is significant at 360 ms ($\beta = 0.768$, $z = 2.13$, $p < 0.05$), and remains significant until convergence at 880 ms after stimulus onset. Factoring in a 200 ms delay to program the eye movement, the results indicate that listeners look to CVNT rather than CVT shortly after onset of coarticulatory nasalization and well before onset of N, that is, vowel nasalization is used to select between the visual options. A similar pattern holds for fixations for voiced [C$\tilde{\text{V}}_{\text{early}}$ND] trials, in which vowel nasalization and N began, on average, 111 ms and 248 ms, respectively, after stimulus onset. For [C$\tilde{\text{V}}_{\text{early}}$ND], CVND-CVD and CVND-CVNT (filled circles versus triangles, upper panel) begin to diverge 360 ms after stimulus onset, with the difference being significant at 380 ms ($\beta = 0.723$, $z = 2.19$, $p < 0.05$). (Because the voiced curves are, over time, less divergent than the voiceless, the differences hover between $p < 0.05$ and $p < 0.15$ for the next 240 ms, until consistent statistical convergence of CVND-CVD and CVND-CVNT at 620 ms after stimulus onset.)

In summary, Hypothesis 1 is largely upheld by the two types of fixation measures. That mean latencies (summarized in Fig. 2) for CVNC-CVC trials with early (80%) vowel nasalization are significantly shorter than latencies for trials that are identical except for later (40%) nasalization indicates that listeners are attending to the coarticulatory information. The time course data (summarized in Fig. 3) mirror this pattern, showing a higher proportion of correct fixations over time for early than late nasalization for CVNC-CVC trials. Importantly, the time course data also more directly link the fixation patterns to coarticulatory nasalization: factoring in the standard programming delay, the proportion fixations on corresponding CVNC-CVC and CVNC-CVNC trials are significantly different well before the onset of N. Moreover, proportion fixations are significantly different for these trials at very nearly the same time point after vowel nasalization onset for [C$\tilde{\text{V}}_{\text{early}}$NT] and [C$\tilde{\text{V}}_{\text{early}}$ND] stimuli, indicating that listeners closely attend to vowel nasalization in both voicing contexts.

A clear exception to these patterns emerged in comparisons with the [C$\tilde{\text{V}}_{\text{late}}$ND] auditory stimuli, for which fixation patterns in the two competitor conditions, CVND-CVD and CVND-CVNT, were unexpectedly the same. Even if we were to conclude that listeners attend less to vowel nasality in voiced contexts, the result is surprising in that listeners should have used N in CVND-CVD (but not CVND-CVNT) trials. (Although [n] might be expected to be less perceptible when followed by glottal pulsing for [d] than when followed by silence for [t], [n] nonetheless should have been informative.) The results for each word show that the latency and proportion correct fixation results look much as predicted for three of the five comparisons. The exceptional stimuli are the late nasalization versions of *send* and *wand* even though results for the early nasalization versions of these words (identical to their "late" counterparts except for vowel nasalization) are as expected. Inspection of the sound files did not offer any clues as to the source of the exceptional pattern for these words.

B. Auditory [CVC] trials

Figure 5 gives the average latencies of the first correct fixations on the trials in which participants heard an oral vowel followed by an oral consonant. A linear mixed model in which Visual Competitor (CVD, CVT, CVNT, CVND) was the fixed effect was fit to first correct fixation latencies for auditory [CVC] stimuli; participant and item were included as random intercepts. If, as we predicted (Hypothesis 2), an oral vowel does not serve as a clearly disambiguating cue even when the competitor image is of a word that would be expected to have a nasalized vowel

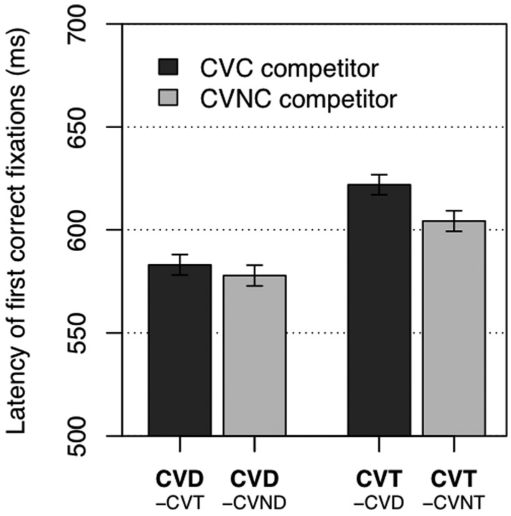

FIG. 5. Mean latency of first correct fixations on trials with auditory [CVC] according to competitor picture (bar type) and coda voicing (voiced: left set of bars; voiceless: right set).

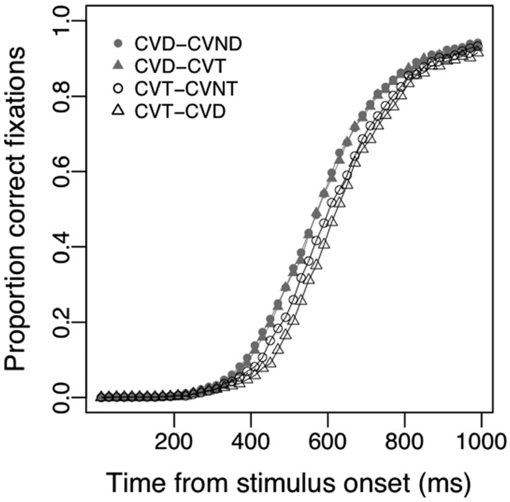

FIG. 6. Pooled proportion correct fixations on trials with auditory [CVC] according to voicing (filled vs unfilled symbols) and competitor image (circles vs triangles).

(CVNC), then latencies for the corresponding CVC-CVNC and CVC-CVC conditions (i.e., paired dark and light bars in Fig. 5) should be the same. The prediction is upheld for the voiced stimuli, where the difference between the two competitor conditions is under 4 ms and the paired comparison (CVD-CVND: CVD-CVT) is not significant ($p > 0.10$). However, when the coda stop was voiceless, latencies were significantly shorter in the CVT-CVNT condition than in the CVT-CVD condition ($\beta = 19.54$, $t = 3.46$, $p < 0.001$)—that is, latencies were shorter in the condition in which the vowels of the imaged words might, under usual coarticulatory conditions, differ in nasality. For all five auditory [CVT] stimuli, initial correct fixations were faster when the competitor image was CVNT than when it was CVD.

Although Fig. 5 shows the latencies to be shorter for the auditory [d]-final than for the corresponding [t]-final stimuli (compare dark to dark and light to light bars), these differences were not significant ($p > 0.10$). We attribute this pattern of results primarily to exceptionally short latencies for auditory [bɛd] (*bed*), for both *bend* and *bet* competitor images. Latencies for auditory [bɛd] were 88 ms shorter than for auditory [bɛt] (*bet*), and 133 ms shorter than the average for the remaining stimuli. Because latencies for auditory [bɛ̃nd], whose initial stop and onset portion of the vowel were acoustically identical to those of [bɛd], were not similarly short, we expect that the exceptional [bɛd] latencies are due not to the auditory stimulus but rather to the visual stimulus: the *bed* image is the most readily identifiable picture in the set. Removing the latencies for auditory [bɛd] and [bɛt] reduces the average difference between the latencies for [d]-final and [t]-final stimuli by half. (Duration and f0 differences between the vowels of CVD and CVT stimuli might have served as potential cues for an upcoming [d], but this should have led to shorter latencies for CVD-CVT than for CVD-CVND trials, which was not the case.)

The proportions of target fixations over time for auditory [CVC] mirror the latency patterns. Figure 6 gives the pooled mean proportion of correct fixations when participants heard [CVD] and [CVT] prompts. A linear mixed model was fit to these fixation proportions using the same time window as for the [CṼNC] stimuli. For the [d]-final stimuli in which participants heard an oral vowel, there was no influence of competitor picture on proportion correct fixations (filled circles versus triangles in Fig. 6; $p > 0.10$). However, for [t]-final stimuli, hearing an oral vowel led to more correct fixations in the condition in which nasalization is disambiguating (unfilled circles versus triangles; $\beta = 0.156$, $z = 6.28$, $p < 0.0001$). That is, as was the case for latency of first correct fixations, the proportion correct fixations is consistent with an oral vowel facilitating lexical decisions in the voiceless context. Thus, results for the voiceless trials disconfirm Hypothesis 2, which predicted no effect of whether the competitor image is of a CVC or CVNC word when listeners hear [CVC].

Another pattern evident in Fig. 6 is that the mean proportion of fixations on the target was overall higher in the auditory [CVD] than [CVT] conditions, although the pattern is again primarily due to especially fast and accurate fixations on visually identifiable *bed*. (When *bed* and *bet* auditory trials are excluded, the proportion fixations on [CVD] trials are very similar to those for the CVT-CVNT condition.) The voicing difference reached significance for comparisons in which the visual competitors were images of CVC words (filled versus unfilled triangles; $\beta = 0.330$, $z = 2.18$, $p < 0.05$) but not when the competitors represented CVNC words (filled versus unfilled circles; $p > 0.10$).

C. Auditory [CṼC] trials

Latencies of correct fixations on auditory [C$\tilde{\text{V}}_{\text{early}}$NC] trials (Sec. III A) showed that participants use vowel nasalization to anticipate a nasal consonant in both CVNC$_{\text{voiceless}}$ and CVNC$_{\text{voiced}}$ contexts. However, in American English, coarticulatory vowel nasalization is often more extensive, and N shorter, before voiceless than before voiced codas. The [CṼC] stimuli, in which N was excised from original [CṼNC] tokens, were included to test whether listeners use these voicing-dependent temporal patterns in making lexical decisions.

For [CṼC] trials, the visual stimuli were always CVNC-CVC and vowel nasalization was [Ṽ$_{\text{early}}$]. Hypothesis 1 predicts that latencies of first fixations of CVNC should be the same for [CṼT] and [CṼD] stimuli. That is, if listeners use coarticulatory nasalization shortly after that information becomes available, first fixations on CVNT and CVND should occur at nearly the same time because the timing of the onset of vowel nasalization is nearly the same in [CṼT] (102 ms) and [CṼD] (111 ms) stimuli. First correct fixations on [CṼC] and corresponding [CṼ$_{\text{early}}$NC] trials should also not differ because these stimuli are identical except for N. However, Hypothesis 3 predicts that the CVNC image should, *over time*, elicit fewer fixations when participants hear [CṼD] than when they hear [CṼT] stimuli because N is longer and more reliably present in voiced than in voiceless contexts. That is, although hearing [CṼ] should elicit looks to the CVND image, hearing [d] (without [n]) may cause listeners to saccade back and forth between the images, or to fixate on the CVD image (or otherwise systematically look away from the CVND image).

1. Pooled results

Figure 7 gives the average latencies of the first correct fixations (where "correct" is taken to be CVNC) on the trials in which participants heard [CṼC] in comparison to the looks to the same image when listeners heard [CṼ$_{\text{early}}$NC]. (The [CṼ$_{\text{early}}$NC] results are the same as those reported in Fig. 2.) A linear mixed model in which N Deletion ([CṼC], [CṼ$_{\text{early}}$NC]) and Visual Competitor (CVD, CVT) were the fixed effects, and participant and item were random intercepts, was computed. In contrast to the latency analyses described above, the cells of this analysis differed substantially in the number of data points available. Each listener heard 20 CṼT and 20 CṼD trials, but not all listeners looked to the image representing a CVNC word on all CṼC trials (and some listeners did so on very few trials, as discussed below). Consequently, there were fewer first correct fixations on the CṼC trials than on the CṼ$_{\text{early}}$NC trials.

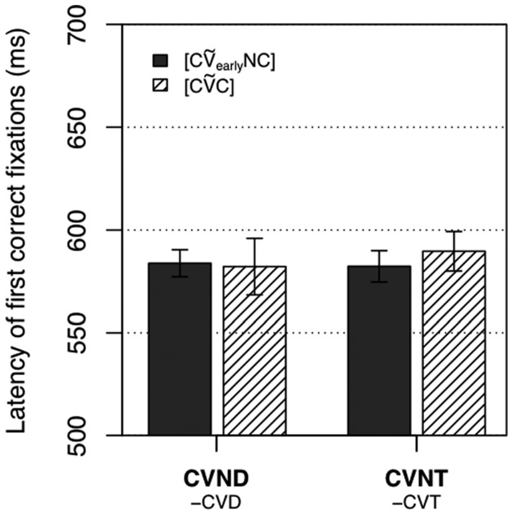

FIG. 7. Mean latency of first correct fixations on CVNC-CVC visual trials according to auditory stimulus (bar type) and coda voicing (voiced: left set of bars; voiceless: right set).

No paired comparison was significant except for the difference between [CṼD] and [CṼ$_{\text{early}}$ND] (left pair of bars in Fig. 7; $\beta = 29.91$, $t = 2.61$, $p < 0.01$). This significant difference, despite nearly identical mean latencies in the two conditions, is due to substantially greater variance in the [CṼD] condition, that is, in the condition with many fewer correct fixations. The source of the [CṼD] variance is inclusion of participant as a random intercept; a revised model with only item as the random variable yields no significant difference for any comparison while the alternative model with only participant as the random variable gives the same pattern as the original model. We return to the across-listener differences in Sec. III C 2.

The pooled mean proportions of correct fixations when participants heard [CṼC] and [CṼ$_{\text{early}}$NC] auditory prompts are given in Fig. 8. As would be expected, in both voicing contexts listeners were more likely to look at the CVNC image when the auditory stimuli included N than when N was deleted. The results of a linear mixed model fit to these fixation proportions (for the same time window as in previous comparisons) showed that participants were significantly

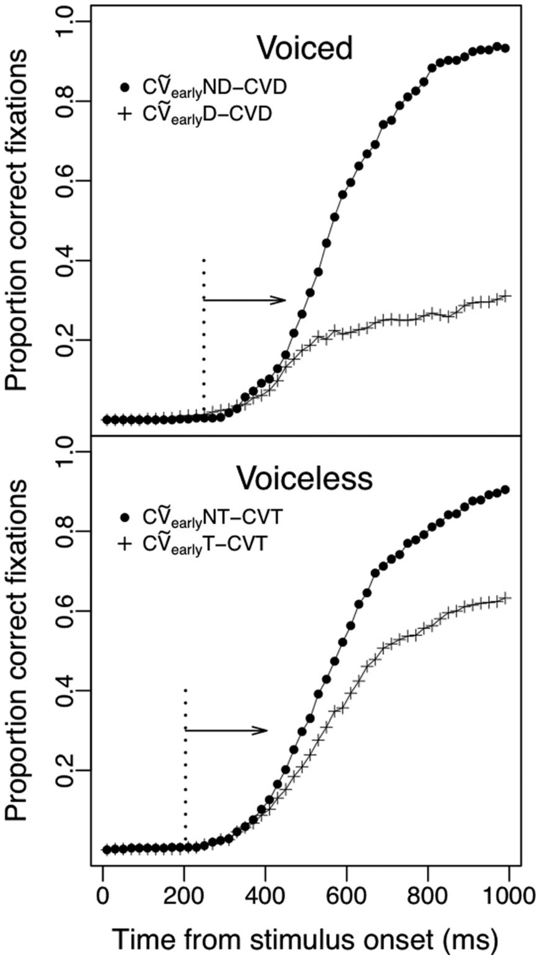

FIG. 8. Pooled proportion correct fixations on trials with auditory [CṼC] (pluses) and [CṼ$_{\text{early}}$NC] (circles) in voiced and voiceless conditions. Dotted lines indicate location of N excision; arrows indicate 200 ms eye movement programming delay.

more likely to look at the CVNC image when they heard $\mathrm{C\tilde{V}_{early}NC}$ than when they heard $\mathrm{C\tilde{V}C}$ in both the voiced ($\beta = 0.776$, $z = 20.39$, $p < 0.0001$) and voiceless ($\beta = 0.335$, $z = 9.75$, $p < 0.0001$) trials. Furthermore, as predicted, the difference between auditory $\mathrm{C\tilde{V}_{early}NC}$ and auditory $\mathrm{C\tilde{V}C}$ was greater when coda C was voiced than when it was voiceless. Thus, although participants were equally likely to look at the CVNC image when they heard $\mathrm{C\tilde{V}_{early}ND}$ and $\mathrm{C\tilde{V}_{early}NT}$ (circles in each panel; $p > 0.10$), when the N was deleted they were more likely to look at the CVNT image than the CVND image (pluses; $\beta = 0.435$, $z = 3.75$, $p < 0.001$). Clearly, vowel nasalization alone, without a nasal consonant, is a more convincing instance of a CVNC word in voiceless than in voiced contexts.

2. *Individual listener results*

Fixation patterns to $[\mathrm{C\tilde{V}C}]$ stimuli differed considerably across listeners, particularly in their responses to $[\mathrm{C\tilde{V}D}]$. This variation, already apparent in the statistical results for the latency data, is graphically represented in Fig. 9 which gives, for each listener, the latency of first correct fixations for the $[\mathrm{C\tilde{V}C}]$ and $[\mathrm{C\tilde{V}_{early}NC}]$ stimuli. Recall that the latencies for the two trial types are not entirely comparable; for example, although $[\mathrm{C\tilde{V}_{early}ND}]$ latencies for all but one listener are based on 19–20 first fixations on the CVND image (out of 20 trials), $[\mathrm{C\tilde{V}D}]$ latencies for several listeners are based on fewer than 10 first fixations. With this caveat, in the voiced condition, listeners whose eye movements to the target had especially long latencies on $[\mathrm{C\tilde{V}_{early}NC}]$ trials tended to have considerably longer latencies on $[\mathrm{C\tilde{V}C}]$ trials. For example, listeners who were slow to look to *bend* when they heard $[\mathrm{b\tilde{\varepsilon}_{early}nd}]$ tended to be even slower to look to the same image when they heard [bɛ̃d]. This is not surprising: listeners who do not reliably use the coarticulatory information to anticipate N (and whose initial correct fixations therefore have relatively long latencies) are not expected to look quickly, if at all, to CVNC when they hear $[\mathrm{C\tilde{V}C}]$. However, the same pattern does not hold for the voiceless context, where latencies of an individual listener's looks on $[\mathrm{C\tilde{V}_{early}NT}]$ and $[\mathrm{C\tilde{V}T}]$ trials are roughly comparable.

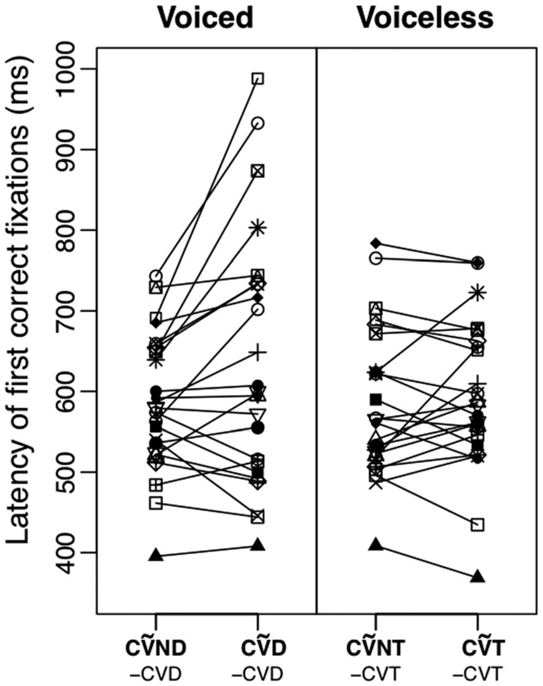

FIG. 9. For each of 23 listeners, latency of first correct fixations on CVNC-CVC visual trials according to coda voicing (left and right panels) and auditory stimulus ($[\mathrm{C\tilde{V}_{early}NC}]$ = left side of each panel and $[\mathrm{C\tilde{V}C}]$ = right side).

Inspection of the time course of correct fixations of individual listeners helps clarify the voicing-dependent patterns that emerge in the latency data. Figure 10 gives the proportion correct fixations for auditory $[\mathrm{C\tilde{V}C}]$ and $[\mathrm{C\tilde{V}_{early}NC}]$ for three individual listeners, each representing a different pattern of response. Listener 221's responses to $[\mathrm{C\tilde{V}C}]$ are broadly representative of those of roughly half of the participants in that the listener attended to vowel nasalization in both voicing conditions, but especially the voiceless. Listener 207 (as well as Listeners 201 and 231; see below) also looked initially to images representing CVND and CVNT words in response to hearing a nasal vowel, but then looked away from CVND. The plateau in this listener's fixations on $[\mathrm{C\tilde{V}D}]$ trials (at around 500 ms) begins approximately 60 ms after the point at which N should have occurred.

Although the majority of listeners looked at least initially to CVNC images in response to auditory $[\mathrm{C\tilde{V}C}]$, for some listeners $[\tilde{\mathrm{V}}]$ did not elicit looks to CVNC in some or all trial types. Listener 212 looked consistently at the CVNT image in response to $[\mathrm{C\tilde{V}T}]$, but at the CVD image in response to $[\mathrm{C\tilde{V}D}]$; four other listeners (Listeners 222, 228, 232, and 233 in Fig. 11) showed a similar response pattern. Two listeners (204 and 214; Fig. 11) looked predominantly at the competitor CVC image in all trial types—voiced and voiceless—with $[\mathrm{C\tilde{V}C}]$.

These individual patterns of responses to $[\mathrm{C\tilde{V}C}]$ trials indicate that listeners differ in the perceptual importance of coarticulatory vowel nasalization in accessing CVNC words, particularly CVND words. To quantify the individual differences and to assess whether they hold for other test conditions, we tested whether a listener's responses to $[\mathrm{C\tilde{V}C}]$ stimuli correlated with that listener's responses to $[\mathrm{C\tilde{V}NC}]$. Listeners who use $[\tilde{\mathrm{V}}]$ in $[\mathrm{C\tilde{V}C}]$ trials should also use the coarticulatory information in $[\mathrm{C\tilde{V}NC}]$ trials, resulting in shorter latencies of first correct fixations in the latter trials. For each listener we calculated, separately for voiced and voiceless trials, the proportion looks to CVNC during the last 40 ms of $[\mathrm{C\tilde{V}C}]$ trials (at which point listeners would have settled on their final response). These values were regressed on the latency of first correct fixations in the $[\mathrm{C\tilde{V}_{early}NC}]$ trials. The prediction is that, the higher the proportion of CVNC fixations in the $[\mathrm{C\tilde{V}C}]$ trials, the shorter the latency of first CVNC fixations in $[\mathrm{C\tilde{V}_{early}NC}]$ trials.

The prediction is upheld for the voiced context, as shown in Fig. 11. The solid regression lines include data from all 23 participants; the dashed lines exclude the results of the three clearest "look away" listeners (Listeners 201, 207, and 231)—that is, listeners who, in $[\mathrm{C\tilde{V}D}]$ trials, initially looked to the CVND item and then looked away (as did Listener 207 in Fig. 10). That R^2 in the voiced condition increases—from 0.11 ($p = 0.12$) to 0.23 [$t(18) = 2.34$,

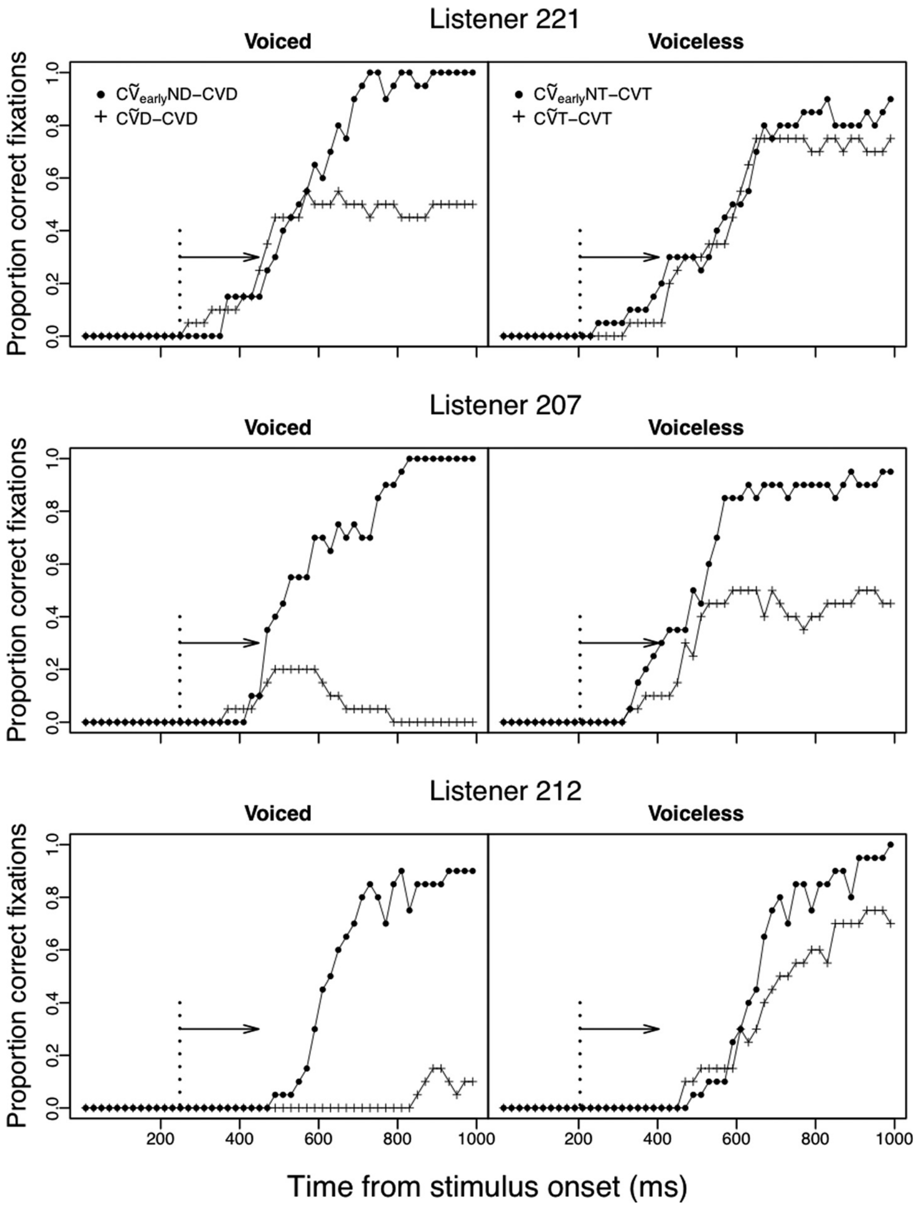

FIG. 10. Proportion correct fixations of three listeners to trials with auditory [CṼC] and [CṼ$_{\text{early}}$NC] in voiced and voiceless conditions.

$p < 0.05$]—when the "look away" listeners are removed is expected because their looks to [CṼD] trials during the final portion of the trial is not representative of their use of [Ṽ]. That is, these listeners initially use vowel nasalization to look to CVND, but [Ṽ] alone is not sufficient to sustain a CVND percept. The absence of a correlation for the voiceless context ($p = 0.35$ for all listeners and $p = 0.54$ for non-"look away" listeners) is not surprising given substantially less variation in this context. Listeners 204 and 214 are outliers who, as observed above, looked at CVC images in response to both [CṼT] and [CṼD]. But the remaining listeners looked predominantly at CVNC images when hearing [CṼT].

D. Frequency of lexical usage

The statistical models reported above tested the three hypotheses. We also tested—independent of our hypotheses about listeners' fixation patterns for the different trial types—whether frequency of lexical usage (Table IV) predicted latencies of first correct fixations. The question was whether, for a given visual pairing, fixation latencies correlated with the lexical frequency of the words corresponding to those images (e.g., whether listeners fixated *went* more quickly than *wet* because *went* is used more often). For each target-visual competitor pairing (for all trial types), we calculated two difference scores: (a) mean first fixation latency of looks to target (e.g., looks to *went* image in a *went-wet* [CṼNC] trial) minus mean first fixation latency of looks to competitor (e.g., looks to *wet* image in a *wet-went* [CVC] trial) and (b) the log frequency difference between target and competitor words (e.g., log frequency$_{\text{went}}$-log frequency$_{\text{wet}}$). A model in which mean fixation latency difference was the independent variable showed that latency differences are not predicted by log frequency ($R^2 = 0.00086$, $t = 0.181$, $p = 0.858$).

IV. DISCUSSION

Listeners' eye movements were monitored as they listened to words with oral vowels or with late or early onset of

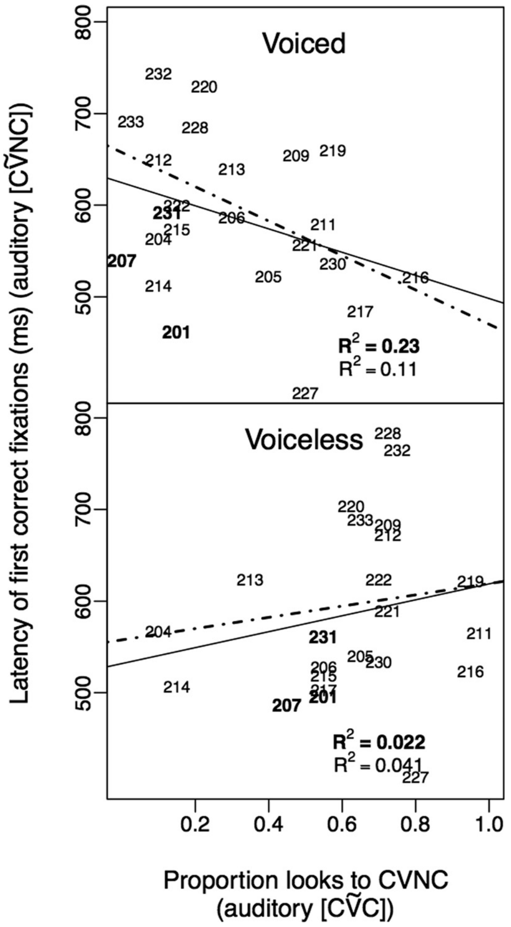

FIG. 11. For voiced and voiceless contexts, scatter plot of relation between proportion CVNC fixations in auditory [CṼC] trials (x axis) against latency of initial correct fixations in auditory [CṼNC] trials (y axis). Numbers: individual listeners. Solid regression lines: all listeners; dashed: "look away" listeners (in bold) excluded. (See text for explanation.)

vowel nasalization. Previous research on perception of coarticulated speech, including perception of anticipatorily nasalized vowels, has shown that coarticulatory information influences listeners' perceptual choices. This study investigated whether listeners use anticipatory velum lowering during a vowel as information about an upcoming consonant as the acoustic signal unfolds over time. We measured fixation patterns on images representing words with and without nasal consonants to test our hypotheses about the use of nasal coarticulation. Two degrees of nasalization were included to assess how closely listeners track the time-varying information: do listeners anticipate a nasal consonant shortly after the acoustic information about velum lowering becomes available?

Overall, participants' fixations on a target as opposed to competitor image indicate that listeners closely attend to the information afforded by coarticulation. When listeners heard a [CṼNC] stimulus, they looked earlier to the image representing a CVNC word on trials in which vowel nasalization would serve as a cue to disambiguate the two images (Fig. 3). When a 200-ms eye movement programming delay is taken into consideration, the proportion correct fixation functions for disambiguating CVNC-CVC and non-disambiguating CVNC-CVNC trials began to diverge visually within about 40 ms of onset of vowel nasalization and statistically within 60 ms (for early vowel nasalization); divergence occurred well before onset of the nasal consonant.

Responses to stimuli that differ in the timing of the onset of nasal coarticulation provide further evidence that listeners track acoustic information for velum lowering in ways that facilitate perception of CVNC words. In trials in which vowel nasalization again provides disambiguating information (i.e., CVNC-CVC), the earlier the coarticulatory information, the faster listeners fixate the target image (Figs. 2 and 3). The advantage afforded by early vocalic information for velum lowering extends throughout the course of the auditory stimulus and beyond (although the early and late nasalization functions do eventually converge).

That most listeners anticipate N as coarticulatory cues become available is also clearly demonstrated in responses to [CṼC] stimuli, in which the nasal consonant has been deleted. The proportion correct fixations over time for the [CṼC] and [CṼNC] stimuli have essentially identical trajectories for the first 440 ms and 460 ms of the voiceless and voiced trials, respectively (Fig. 8). These identical time courses indicate that, up to this point in the unfolding acoustic signal, fixations on images of CVNC words were based on vowel nasalization. Again factoring in eye movement delay, differences between fixations on [CṼC] and [CṼNC] trials begin to occur about 20–25 ms after the point of N deletion in [CṼC] stimuli. Thus, as acoustic information about the velum lowering gesture becomes available over the course of a VNC sequence, listeners use this information in ways that result in faster and, at least temporarily, more accurate perception of the target word.

This study also investigated whether an oral vowel provides similarly facilitative information about VC sequences. We predicted (Hypothesis 2) that it would not because a vowel that is oral throughout much of its articulation is not incompatible with an upcoming N. Moreover, due to perceptual compensation, even articulatorily nasalized vowels are perceived as relatively oral when followed by N. Our prediction was only partially upheld: an oral vowel did not shorten fixation latencies or increase correct fixations when followed by a voiced obstruent, but it did when the coda was voiceless. For the fixation latencies, the advantage in the CVT-CVNT condition relative to the CVT-CVD condition was small—7 to 13 ms—for the prompts *bet*, *set*, *wet*, and *let*, but was roughly 50 ms for *watt*. The larger effect for *watt* might not be entirely due to orality. For the Michigan speaker in this experiment, and quite possibly for many of our mostly Midwestern listeners, the vowel in *want* has slightly more lip rounding than the vowels in *watt*, *wad*, or *wand*. This difference may have provided nasality-independent information about the target item for visual *watt*-*want*—information that was not available in the corresponding voiced *wad*-*wand* comparison. Nonetheless, when listeners' visual options were CVC-CVNC as opposed to CVC-CVC, hearing an oral vowel in the disambiguating condition provided listeners with a small but consistent advantage in the voiceless but not the voiced context. That is, presenting listeners with an oral vowel and showing them competing images whose vowels would normally differ in nasality speeds correct fixations, and increases correct looks to the target, only in the context in which the vowel of the CVNC item would be especially heavily nasalized and N would be short or even absent. In the (voiced) context in which the competitor CVNC would often have somewhat later [$\tilde{\text{V}}$] onset and a reliably present N, an oral vowel did not facilitate looks to the CVC item.

A robust outcome of this study is that, although listeners attend closely to the information for changing vocal tract configuration in their moment-by-moment processing, they are not simply responding to the coarticulatory cues shortly after they become available. Rather, their response is selective in at least two respects. First, as a group, listeners' attention to coarticulatory information is context-dependent: they are more likely to use, or assign a heavier perceptual weight to, coarticulatory nasalization in the phonetic context in which it might be especially important (Hypothesis 3). Second, individual listeners are differently selective in their perceptual weights.

We interpret first the effects of voicing context on the time course of perception of anticipatory nasalization. When participants are provided with early coarticulatory information for an upcoming nasal consonant, the initial perceptual time course is nearly identical for CVNT and CVND words. That is, in response to [C$\tilde{\text{V}}_{\text{early}}$], eye movements have the same latencies on CVNT-CVT and CVND-CVD trials (Fig. 2), and the time point at which fixations on disambiguating CVNC-CVC trials diverge from non-disambiguating CVNC-CVNC is the same for the voiceless and voiced contexts (relative to nasalization onset; Fig. 3). Thus, early, clear acoustic information for velum lowering is initially used by listeners to access CVNC words independent of voicing context.

However, as the input signal continues to unfold, voicing effects begin to emerge even when strong coarticulatory cues for an upcoming nasal rather than oral consonant are available. For example, as the acoustic signal for [C$\tilde{\text{V}}_{\text{early}}$NC] unfolds over time, a bin-by-bin comparison of the proportion looks to CVNC shows a greater statistical divergence between the disambiguating and non-disambiguating conditions in the voiceless than in the voiced context. This voicing difference is particularly striking in that, in the non-disambiguating context, acoustic information that the final C is [t] or [d] is available, on average, approximately 65 ms earlier for [t] due to shorter V and N durations, which might be expected to lead to earlier convergence for the voiceless context. Thus, although listeners are equally likely to look initially to images representing CVNT and CVND words based on early coarticulatory vowel nasalization, over time the anticipatory information affords a greater perceptual advantage in the voiceless context.

Voicing differences emerge more clearly when the coarticulatory cues become available later in the acoustic input—either through late onset of vowel nasalization or an oral vowel. In both cases, as the initial portion of the vowel unfolds, no coarticulatory information is available to help listeners select between CVNC and CVC. When the vowel remains oral ([CVC]), non-nasality leads to increased looks to the correct CVC image, but only in the voiceless context (Figs. 5 and 6). In [C$\tilde{\text{V}}_{\text{late}}$NC] utterances, acoustic information for velum lowering begins after 60% of the vowel has occurred. In the voiceless context, this information is sufficient for listeners to look earlier and more often to CVNC than CVC images (Figs. 2 and 3). A perceptual advantage afforded by late cues for velum lowering is not found, though, in the voiced context. The difference cannot be attributed to naturalness of the coarticulatory patterns because, as previous studies have shown, later onset of vowel nasalization is more likely before voiced codas (e.g., Malécot, 1960; Cohn, 1990). Rather, for both [CVC] and [C$\tilde{\text{V}}_{\text{late}}$NC] inputs, we interpret the patterns of eye movements to target images as indicating that listeners are especially sensitive to evolving acoustic information about velum lowering (or absence of velum lowering) in the phonetic context in which that information is especially robust—and in which the information is especially important, given that [$\tilde{\text{V}}$] may be the only source of information for velum lowering in the articulatory realization of CVNT words by some speakers. Malécot (1960) and, more recently, Beddor (2009) have shown that American English listeners weighted vowel nasalization more heavily in voiceless than in voiced contexts in their identification judgments [see also Treiman *et al.* (1995) for data from children]. The eye tracking patterns suggest that these perceptual biases influence not only listeners' final lexical decisions, but also their attentiveness to coarticulation in their moment-by-moment processing of that information.

Eye movements in response to auditory stimuli in which the nasal consonant was absent further delineate the nature of context-dependent processing of coarticulation in real time. Parallel to the [C$\tilde{\text{V}}_{\text{early}}$NC] condition, when listeners heard [C$\tilde{\text{V}}$C] stimuli (in which nasalization onset also began early in the vowel), they were initially equally likely to look to CVNC in voiced and voiceless contexts on the basis of anticipatory coarticulation. As information for coda C became available, [C$\tilde{\text{V}}$T] continued to elicit increased looks to CVNT over time for nearly all listeners. [C$\tilde{\text{V}}$D] elicited much more variable responses (see below), although the overall result was that, by the end of the trial, the competitor image CVD elicited nearly twice as many looks as did CVND.

The nature and source of individual differences in the processing of coarticulatory information have theoretical implications. We have discussed elsewhere the implications of listener-specific perception of coarticulated speech, including listener differences in "offline" identification and discrimination, for theories of sound change (Beddor, 2009, 2012). Of particular interest here are the implications of real time processing differences for theories of speech perception. As has been shown, there is greater across-listener variation in eye movements in response to stimuli with voiced than with voiceless codas. In [C$\tilde{\text{V}}$D] trials, for example, expectations about an upcoming N on the basis of [$\tilde{\text{V}}$] are maintained, despite the lack of N, for some but not all listeners. For some listeners in the latter group, the perceptual uncertainty led to a leveling off of looks to the CVND image whereas other listeners revised their initial looks to fixate consistently on CVD. [This "look away" response is analogous to recovery from a "garden path" in a syntactically ambiguous sentence, e.g., Tanenhaus *et al.* (1995).] For yet other listeners, early onset of vowel nasalization in [CṼND] and [CṼD] was not sufficient to elicit looks to CVND; for them, N was required. (These are, for example, Listener 212 in Fig. 10 and, most extremely, Listeners 212, 222, 232, and 233 in Fig. 11.) It is not surprising that a comparable range of responses to [CṼNT] and [CṼT] was not found. Put simply, for listeners of American English, requiring N before voiceless codas is not a safe bet because the velum lowering gesture does not consistently overlap with the consonantal constriction (Malécot, 1960; Cohn, 1990; Beddor, 2009).

What, then, is the source of across-participant variation in the weight accorded coarticulatory information while processing [CṼND] and [CṼD] inputs? In many respects, an exemplar approach to speech perception is ideally suited to handling listener variation, including listener differences in weights assigned to particular stored exemplars or to particular signal properties when categorizing a new input (Johnson, 1997). Specifically, exemplars of frequent, recent experiences are heavily weighted (i.e., have a high activation level) and especially influence categorization of a new token (Pierrehumbert, 2001). For the real-time processing data reported here, an individual listener's experiences with words of the structure CVND (and possibly CVD, CVNT) could perhaps shape the weights that listener assigns to [Ṽ] and N when hearing [CṼND] as it unfolds over time. Information on participants' linguistic background might be expected to reveal the predicted experiential source, although a listener's exemplars are not solely determined by raw experience. Rather, exemplars to which a listener pays greater attention are expected to have a correspondingly greater impact on the resulting exemplar space (Johnson, 1997; Pierrehumbert, 2001). An elaboration of an exemplar model that both specified a detailed array of acoustic features and allowed for feature-specific attentional weights could presumably accommodate much of the across-participant variation observed in this study.

However, we expect that an experiential factor *common* to all participants plays the predominant role in the variation observed in this study. Although all participants would have had experience with substantial variation in the temporal extent of vowel nasalization and the duration of the nasal consonant, both [Ṽ] and N would likely have been reliably present in most CVND words that they have heard. Some listeners may attend to [Ṽ] while others attend more to N precisely because both vocalic and consonantal indicators of velum lowering are available when the input is a CVND word: listeners have multiple cues available to them for CVND. [For CVNT words, the situation is different in that American English listeners have heard utterances with [Ṽ] but lacking N, rendering the vowel the single consistent indicator of velum lowering. See also Toscano and McMurray (2010) for discussion of weighting of cues as a function of their reliability.] The choices for CVND words of how to weight multiple cues influence the processing of these words over time. A gesturalist approach, or more generally an approach that emphasizes the perceptual value of coarticulated signals, leads us to expect most outcomes of this study. When coarticulatory information for velum lowering is available early in the signal, most listeners use that information soon after it occurs to anticipate an upcoming nasal consonant. In CVND contexts, that vowel nasalization needs to be reinforced by N for some listeners can be accounted for by perceptual attunement to the coarticulatory patterns of English (e.g., Best, 1995). Unexpected within these approaches, though, is the small minority of listeners who do not use the coarticulatory information in the voiced context, despite its potential to disambiguate the target and competitor items.

In summary, listeners attend to the acoustic effects of overlapping articulations in real-time processing. As the signal unfolds, listeners' moment-by-moment fixations on visual displays indicate that they are actively using the emerging coarticulatory information to select a target image over its (acoustically and articulatorily) minimally distinct competitor. Moreover, especially when coarticulatory cues are available early in the input signal, as was the case for [CṼ$_{\text{early}}$NC] and [CṼC] stimuli, the time course of listeners' eye movements to target images indicates that listeners use these disambiguating cues shortly after they become available. However, even for a given gesture, such as velum lowering, not all coarticulatory information is equally useful or accorded equal attention. Rather, processing of coarticulation is partially dependent on the not-yet-heard but expected (based on visual response options) phonetic context. The context-dependent perceptual patterns are fully consistent with the articulatory timing of velum lowering in different voicing contexts. Presumably, knowledge of the detailed timing of coarticulation influences the perceptual weight assigned to the anticipatory cues. These perceptual weights are listener-specific, at least in the phonetic context in which both the coarticulatory cue and its source (here, N) are consistently realized. At this stage of our research, we do not know whether the strong contextual and individual listener patterns in the processing of unfolding coarticulatory cues are specific to vowel nasalization, or perhaps to relatively long-distance coarticulatory effects. Regardless, the time course of perception of velum lowering in American English indicates that the dynamics of perception parallel the dynamics of the gestural information encoded in the acoustic signal. In real-time processing, listeners closely track coarticulatory dynamics in ways that speed lexical activation.

ACKNOWLEDGMENTS

We thank Matthew Masapollo, Alan Mishler, Cortney Sera, and Damon Tutunjian for assistance with data collection. This work benefited from an early suggestion by Taehong Cho and from the comments of members of several audiences, especially those at Nasal 2009 (Montpellier, France) and Laboratory Phonology 12 (Albuquerque, NM), and the University of Michigan Phonetics-Phonology Research Group. We are grateful as well to two anonymous reviewers for their valuable suggestions.

Alfonso, P. J., and Baer, T. (**1982**). "Dynamics of vowel articulation," Lang. Speech **25**, 151–173.

Allopenna. P. D., Magnuson, J. S., and Tanenhaus, M. K. (**1998**). "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," J. Mem. Lang. **38**, 419–439.

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge, UK), Chap. 7, pp. 241–302.

Bates, D., Maechler, M., and Bolker, B. (**2011**). "lme4: Linear mixed-effects models using S4 classes, R package version 0.999375-42," http://CRAN.R-project.org/package=lme4 (Last viewed April 30, 2012).

Beddor, P. S. (**2009**). "A coarticulatory path to sound change," Language **85**, 785–821.

Beddor, P. S. (**2012**). "Perception grammars and sound change," in *The Initiation of Sound Change: Production, Perception, and Social Factors*, edited by M.-J. Solé and D. Recasens (John Benjamins, Amsterdam), pp. 37–55.

Beddor, P. S., and Krakow, R. A. (**1999**). "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," J. Acoust. Soc. Am. **106**, 2868–2887.

Bell-Berti, F. (**1993**). "Understanding velic motor control: Studies of segmental context," in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic Press, New York), pp. 63–85.

Bell-Berti, F., Baer, T., Harris, K. S., and Niimi, S. (**1979**). "Coarticulatory effects of vowel quality on velar function," Phonetica **36**, 187–193.

Best, C. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 171–204.

Bradlow, A. R. (**2002**). "Confluent talker-and listener-oriented forces in clear speech production," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, New York), pp. 241–273.

Boersma, P., and Weenink, D. (**2009**). Praat: Doing phonetics by computer [Computer program], http://www.praat.org/ (Last viewed May 18, 2009).

Clumeck, H. (**1976**). "Patterns of soft palate movements in six languages," J. Phonetics **4**, 337–351.

Cohn, A. C. (**1990**). "Phonetic and phonological rules of nasalization," *UCLA Working Pap. Phonetics* 76, 1–224.

Connine, C. M., and Darnieder, L. M. (**2009**). "Perceptual learning of co-articulation in speech," J. Mem. Lang. **61**, 368–378.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (**2001**). "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition," Lang. Cogn. Process. **16**, 507–534.

Dahan, D., and Tanenhaus, M. K. (**2004**). "Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints," J. Exp. Psychol.: Learning Memory Cogn. **30**, 498–513.

Davies, M. (**2008**). "The corpus of contemporary American English: 450 million words, 1990–present," http://corpus.byu.edu/coca/ (Last viewed January 19, 2013).

Elman, J. L., and McClelland, J. L. (**1986**). "Exploiting the lawful variability in the speech wave," in *Invariance and Variability of Speech Processes*, edited by J. S. Perkell, and D. H. Klatt (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 360–380.

Flagg, E. J., Oram Cardy, J. E., and Roberts, T. P. L. (**2006**). "MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs," Neurosci. Lett. **397**, 263–268.

Fowler, C. A. (**1996**). "Listeners do hear sounds, not tongues," J. Acoust. Soc. Am. **99**, 1730–1741.

Fowler, C. A., and Brown, J. M. (**2000**). "Perceptual parsing of acoustic consequences of velum lowering from information for vowels," Percept. Psychophys. **62**, 21–32.

Gow, D. W., and McMurray, B. (**2007**). "Word recognition and phonology: The case of English coronal place assimilation," in *Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, Berlin), pp. 173–200.

Hawkins, S. (**2003**). "Roles and representations of systematic fine phonetic detail in speech understanding," J. Phon. **31**, 373–405.

Jenkins, J. J., Strange, W., and Trent, S. A. (**1999**). "Context-independent dynamic information for the perception of coarticulated vowels," J. Acoust. Soc. Am. **106**, 438–448.

Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego, CA), pp. 145–166.

Kawasaki, H. (**1986**). "Phonetic explanation for phonological universals: The case of distinctive vowel nasalization," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic Press, Orlando, FL), pp. 81–103.

Krakow, R. A. (**1993**). "Nonsegmental influences in velum movement patterns: Syllables, sentences, stress, and speaking rate," in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic Press, New York), pp. 87–113.

Krakow, R. A. (**1999**). "Physiological organization of syllables: A review," J. Phonetics **27**, 23–54.

Lahiri, A., and Marslen-Wilson, W. (**1991**). "The mental representation of lexical form: A phonological approach to the recognition lexicon," Cognition **38**, 245–294.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht, Netherlands), pp. 403–439.

Malécot, A. (**1960**). "Vowel nasality as a distinctive feature in American English," Language **36**, 222–229.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.

Martin, J. G., and Bunnell, H. T. (**1981**). "Perception of anticipatory coarticulation effects," J. Acoust. Soc. Am. **69**, 559–567.

Matthies, M., Perrier, P., Perkell, J. S., and Zandipour, M. (**2001**). "Variation in anticipatory coarticulation with changes in clarity and rate," J. Speech Language Hear. Res. **44**, 340–353.

Moll, K. L. "Velopharyngeal closure on vowels," J. Speech Hear. Res. **5**, 30–37 (1962).

Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.

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

Ohala, J. J., and Ohala, M. (**1995**). "Speech perception and lexical representation: The role of vowel nasalization in Hindi and English," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge, UK), pp. 41–60.

Ostreicher, H. J., and Sharf, D. J. (**1976**). "Effects of coarticulation on the identification of deleted consonant and vowel sounds," J. Phonetics **4**, 285–301.

Pierrehumbert, J. B. (**2001**). "Exemplar dynamics: Word frequency, lenition and contrast," in *Frequency Effects and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (John Benjamins, Amsterdam), pp. 137–157.

Raphael, L. J., Dorman, M. F., Freeman, F., and Tobin, C. (**1975**). "Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies," J. Speech Hear. Res. **18**, 389–400.

Scarborough, R. (**2004**). "Coarticulation and the structure of the lexicon," Doctoral dissertation, University of California, Los Angeles, 154 pp.

Solé, M.-J. (**1995**). "Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization," Lang. Speech **38**, 1–23.

Stevens, K. N., and Keyser, S. J. (**2010**). "Quantal theory, enhancement and overlap," J. Phonetics **38**, 10–19.

Strange, W. (**1989**). "Evolving theories of vowel perception," J. Acoust. Soc. Am. **85**, 2081–2087.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (**1995**). "Integration of visual and linguistic information in spoken language comprehension," Science **268**, 1632–1634.

Tatham, M., and Morton, K. (**2006**). *Speech Production and Perception* (Palgrave, New York), Chap. 3, pp. 40–98.

Toscano, J. C., and McMurray, B. (**2010**). "Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributed statistics," Cogn. Sci. **34**, 434–464.

Treiman, R., Zurowski, A., and Richmond-Welty, E. D. (**1995**). "What happened to the 'n' of *sink*? Children's spellings of final consonant clusters," Cognition **55**, 1–38.

Vaissière, J. (**1988**). "Prediction of articulatory movement of the velum from phonetic input," Phonetica **45**, 122–139.

Warren, P., and Marslen-Wilson, W. (**1987**). "Continuous uptake of acoustic cues in spoken word recognition," Percept. Psychophys. **41**, 262–275.

Whalen, D. H. (**1984**). "Subcategorical phonetic mismatches slow phonetic judgments," Percept. Psychophys. **35**, 49–64.

Whalen, D. H. (**1991**). "Subcategorical phonetic mismatches and lexical access," Percept. Psychophys. **50**, 351–360.