Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 37 (2009) 97–110
Journal of Phonetics
www.elsevier.com/locate/phonetics

Timing effects of syllable structure and stress on nasals: A real-time MRI examination

Dani Byrd[a,*], Stephen Tobin[a,1], Erik Bresch[b], Shrikanth Narayanan[a,b]

[a]*Department of Linguistics, USC, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, USA*
[b]*Viterbi School of Engineering, Department of Electrical Engineering, USC, USA*

Received 12 July 2007; received in revised form 29 September 2008; accepted 10 October 2008

Abstract

The coordination of velum and oral gestures for English [n] is studied using real-time magnetic resonance imaging (MRI) movies to reconstruct vocal tract aperture functions. This technique allows for the examination of parts of the vocal tract otherwise inaccessible to dynamic imaging or movement tracking. The present experiment considers syllable onset, coda, and juncture geminate nasals and also addresses the effects of a variety of word stress patterns on segment internal coordination. We find a bimodal timing pattern in which near-synchrony of velum lowering and tongue tip raising characterizes the timing for onsets and temporal lag between the gestures is characteristic for codas, supporting and extending the findings for [m] of Krakow [(1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Doctoral Dissertation, Yale University, New Haven, CT; (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In M. A. Huffman, R. A. Krakow (Eds.), *Nasals, nasalization and the velum (phonetics and phonology V)* (pp. 87–116). New York: Academic Press]. Intervocalic word-internal nasals are found to have timing patterns that are sensitive to the local stress context, which suggests the presence of an underlying timing specification that can yield flexibly. We consider these findings in light of the gestural coupling structures described by Goldstein and colleagues [Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action in units understanding the evolution of phonology. In M. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press; Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2008). Coupled oscillator planning model of speech timing and syllable structure. In *Proceedings of the 8th phonetics conference of China and the international symposium on phonetic frontiers*; Nam, H., Goldstein, L., & Saltzman, E. (in press). Self-organization of syllable structure: A coupled oscillator model. In Chitoran, Coupe, Marsico, & Pellegrino (Eds.), *Approaches to phonological complexity*].
© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Within a view of speech production that considers the units of execution or action to be articulatory gestures, there have been proposals that gestures *cohere* into larger units including (but not limited to) segments (Byrd, 1996a; Saltzman & Byrd, 2000; Saltzman, Lofqvist, & Mitra, 2000; Saltzman & Munhall, 1989). In large measure, this coherence has been attributed to relative timing characteristics among the component gestures. In this study, we will be considering coordination in one instance of a multi-gesture segment in English—the nasal consonant. Nasal stop segments can be understood to include a velum lowering gesture and an oral closure gesture, in English, either bilabial, tongue tip, or tongue body. Specifically, we are interested in examining the intergestural coordination between these two components as a function of syllable structure and stress. This coordination has been difficult to study in the past as an examination of velum movement has been very challenging and has required rather invasive techniques. We will present real-time magnetic resonance imaging (MRI) data (Bresch, Adams et al., 2006; Bresch, Nielsen, Nayak, & Narayanan, 2006; Narayanan, Nayak, Lee, Sethy, & Byrd, 2004; *sail.usc.edu/span*) that allow a view of the vocal tract in its entirety, including the velum, with sufficient temporal resolution to reflect intergestural

*Corresponding author. Tel.: +1 213 821 1227; fax: +1 213 740 9306.
E-mail address: dbyrd@usc.edu (D. Byrd).
[1]Now at the University of Connecticut, Department of Psychology.

0095-4470/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2008.10.002

Author's personal copy

timing patterns. We have an interest in replicating older findings on syllable structure with this new technique and in generating new results on the effects of both syllable structure and stress.

Syllable position effects on spatial magnitude or durational properties of the articulation of *single* gestures have been reported in a variety of studies (e.g., Browman & Goldstein, 1995a, b; Byrd, 1996b; Fougeron & Keating, 1995; Fromkin, 1965; Keating, 1995, and many others). With respect to English nasal stops, the general consensus in the literature is that there is a lower velum posture in word-final position and that this velum lowering occurs during the acoustic interval of the preceding vowel, resulting in nasalization of that vowel (e.g., the acoustic studies of Cohn, 1990; Fujimura & Lovins, 1978; Vassière, 1988); we review the articulatory work below.

Although a number of production studies have examined the influence of syllable position on single gestures, only a few studies have examined the effect of syllable position on timing between gestures composing multi-gesture segments. These studies have shown that timing differences do exist between the component gestures constituting single segments. Specifically, the intergestural timing for English [m], [l], [r], and [w] as a function of syllable position has been investigated (Delattre, 1971; Gick & Goldstein, 2002; Krakow, 1989, 1993; Sproat & Fujimura, 1993).

In her seminal work on velum–oral gestural timing in nasals, Krakow (1989, 1993) examined [m] production for two speakers using a Velotrace (Horiguchi & Bell-Berti, 1987) to transduce velum height changes and Selspot for lip point-tracking. She compared the relative timing of the achievement of maximal velum lowering and the achievement of the bilabial closure gesture. Krakow found a consistent difference between syllable conditions. In the syllable-initial nasals the achievement of maximal velum lowering closely coincided with the achievement of the bilabial closure gesture. In the syllable-final nasals, on the other hand, the end of velum lowering closely coincided with the beginning of lip raising. Thus, Krakow's work suggests that in syllable-initial position the oral and nasal gestures' target achievement are coordinated in a roughly synchronous relationship, while in syllable-final position the velum lowering gesture consistently precedes the oral closure. Similarly, Clumeck (1976) used a Nasograph (Ohala, 1971; which measures light from below the velum reaching a sensor in the nasal cavity) to show proportionally earlier velum lowering in syllable-final nasals in Portuguese. Additionally, work by Krakow and others has also found spatial differences in velum posture or muscular activity as a function of position within the syllable (Krakow, 1999 cites Fujimura, 1990; Kiritani, Hirose, & Sawashima, 1980); generally lower velums are found for final nasals. Our primary interest here is on the syllable-related timing effects.

Looking at the segment internal timing of another two-gesture segment [l], Sproat and Fujimura (1993) considered movement data from five speakers' productions. They found that in word-initial [l]s the tongue tip gesture target leads the tongue body gesture target, while in word-final position the reverse temporal pattern obtains. Similar observations were made from X-ray data in 1971 by Delattre and can also be found in Giles and Moll (1975) and Browman and Goldstein (1995a). For another multi-gesture segment, [r], which has labial, tongue body, and tongue root components, Gick and Goldstein (2002) report that prevocalically, the lip gesture target precedes the tongue body gesture target, which in turn precedes the tongue root gesture target. Postvocalically, the lip and root gestures were found to be smaller and the tongue root gesture target was observed to precede rather than follow the others. Huffman's (1997) acoustic work suggests that stress is also of importance in the intergestural timing of laterals, though the articulatory studies have not examined this.

Strikingly, there appear general parallels (see also Browman & Goldstein, 1992; Krakow, 1999) in the effect of syllable structure on nasals, glides, and liquids—namely, that in onset position the primary oral constriction gesture target precedes or is synchronous with the secondary (non-primary constriction) gesture target (i.e., the pharyngeal gesture for [r], dorsal for [l], or velum for nasals), whereas in coda, the secondary gesture target occurs far earlier, during the preceding vocalic nucleus (see the review in Krakow, 1999 of spatial as well as temporal differences).

Works of Goldstein and colleagues (Goldstein, Byrd, & Saltzman, 2006; Goldstein, Nam, Saltzman, & Chitoran, 2008; Nam, Goldstein, & Saltzman, in press) explain the differences in the articulations of syllable-initial and syllable-final consonants in terms of a broader generalization, namely, that gestural relations in onsets and codas reflect two intrinsically stable modes of coordinating actions: in-phase and antiphase. These two modes are claimed to be found universally in human languages (and in human motor behavior, generally), whereas other more specialized coordination patterns must be learned language-specifically. In the in-phase coupling mode, which is the most stable mode (Haken, Kelso, & Bunz, 1985), gestures are involved in an underlying coordination pattern of synchrony; while in the antiphase coupling mode, the gestures are involved in an underlying coordination pattern that is sequential. Goldstein and colleagues propose that in-phase intergestural coupling relations form the basis for syllable onset structure, and antiphase intergestural coupling relations form the basis of syllable coda structure. "Onset consonant gestures are hypothesized to be coupled in-phase to the tautosyllabic vowel (regardless of how many there are in an onset), while coda consonant gestures are coupled in an antiphase pattern. This topology can account simultaneously for regularities in relative timing and [timing] variability" (Goldstein et al., 2008). Thus, they claim that intrinsically available coordination patterns are the fundamental source of syllable structure cross-linguistically. It is important to note, however, that because gestures can be involved in multiple coordination

Author's personal copy

relationships such as vowel-to-consonant and consonant-to-consonant (Browman & Goldstein, 2000; Goldstein et al., 2008; Nam et al., in press), the observed output timing is the result of numerous interacting forces on underlying temporal relations.

The present study has two motivations in addition to the introduction of a new methodology. Articulatory studies of a number of multi-gesture segments have indicated that there are distinct intergestural timing patterns associated syllable-initial and syllable-final positions, and we anticipate replicating these findings for [n], which has not been examined previously. Krakow's earlier study used a rather invasive technique (the Velotrace), was limited to [m] (omitting lingual nasals due to the use of Selspot movement tracking), and did not illuminate situations in which the syllable structure is more complex, such as juncture geminates, or more ambiguous in terms of stress pattern variations. Our study of [n] will extend previous findings to the only lingual nasal that occurs in syllable onset and coda in English, and our consideration of juncture geminates and stress will substantially extend our knowledge of within-segment multi-gesture coordination.

Juncture geminates refer to the possibility of having an abutting coda and onset consonant at a word boundary, such as "hip pocket" or "bomb man." At least in English, such phonological sequences are typically realized with a single constriction formation and release in casual speech (see e.g., Byrd, 1995a; Byrd, Campos-Astorkiza, & Shepherd, 2006; Munhall & Lofqvist, 1992; Nolan, 1992; see also Browman & Goldstein, 1995a). Primarily, single gesture characteristics have been examined in these few production studies. Multi-gesture coordination within segments has not been evaluated for such sequences. We will examine this for juncture geminate nasals, thereby extending the type of stimuli considered in Krakow's earlier seminal study. Since phonologically such juncture geminate sequences are comprised of a coda followed by an onset, it is unknown whether the intergestural characteristics will exhibit coda-like or onset-like timing or some intermediate pattern. However, we *prima facie* expect the lag between the left edges of the two gestures in the juncture geminate to behave in a coda-like way simply because the initial (leftmost) constriction formations are under examination for this measure and the leftmost constriction forms part of an underlying coda. Further, we will be able to consider whether the velum lowering gesture itself in the juncture geminate has spatial and temporal characteristics like codas or onsets or intermediate in some way. In summary, the first study below, referred to as the syllable condition set, will examine the intergestural velum–oral timing for nasals in coda, onset, and juncture geminate positions using the non-invasive technique of real-time MRI.

The second study below, referred to as the stress condition set, will examine the same intergestural timing relations as in the syllable condition; however, in this case we will turn our attention to word-internal intervocalic nasals (VnV) that are nominally in a syllable onset position (e.g., due to the Maximal Onset Principle) but have been placed with varying preceding and following syllable lexical stress conditions: Unstressed–Stressed, Stressed–Unstressed, and Primary Stressed–Secondary Stressed, and for comparison from stimuli set one a word-initial onset in a Stressed–Stressed context. The syllable affiliation of word-internal intervocalic consonants in English is viewed as ambiguous, particularly when the upcoming syllable does not carry primary stress (Kahn, 1976; also Blevins, 1995).

Part two below will examine whether differential internal organization for such intervocalic nasals arises as a consequence of the stress environment (e.g., falling or rising stress contours). Effects of stress environment on the temporal characteristics of single gestures and their coordination with adjacent sounds have been reported, with flapping in English often being presented as an example of such effects (de Jong, 1998; Fukaya & Byrd, 2003; Stone & Hamlet, 1982). For example, de Jong (1998) suggests that tongue tip gestures become more overlapped with adjacent vowel gestures in flapping environments (falling or sometimes level stress) and that a possible acoustic consequence of this is a perceived flap. Single alveolar segments in this falling-stress environment are typically considered to be very coda-like, i.e., subject to lenition. Turk (1994) reports kinematic data on syllabically ambiguous labial segments finding that their articulatory properties are basically coda-like, not intermediate between a coda and onset.[2]

With regard to multi-gesture segments, Krakow (1993) finds that velum behavior in words like "homey" and "seamy" show a similar timing pattern as that seen for codas in phrases like "home E" and "seam E," while intervocalic [m]s preceding a primary stress like "pomade" behave like onsets as in "pa made." She concludes that the primary stress syllabically *attracts* the nasal and governs the resulting timing pattern. She, like Turk (1994), does not find an intermediate pattern for potentially "ambisyllabic" [m]s but rather finds that they behave like codas. Gick (2003a) further investigated the hypothesis that the pattern of intrasegmental timing in ambisyllabic consonants may be intermediate between that of consonants with unambiguous syllable affiliation. Gick (2003a) predicted that the gestures of such consonants would be critically phased in relation to both the preceding vowel and the following vowel. He claims that such an intergestural phasing relation would result in gestural magnitude and gestural timing intermediate between that observed for onsets and

[2]In her dissertation (Turk, 1993), results were consistent with Turk 's (1994) analyses in that the ambiguous segments were clearly more coda-like than onset-like, but there were some indications in the data of differences between ambiguous consonant articulations and those of clear codas. These differences varied across articulators (sometimes ambiguous consonants formed a clear third category, sometimes some patterned like onsets, some like codas). While difficult to interpret, these results might argue in favor of some type of ambisyllabic analysis.

Author's personal copy

codas. In an experiment on English liquids and glides using magnetometer point-tracking of articulation, Gick found that while unambiguous onsets differed from both unambiguous codas and ambisyllabic consonants in their temporal patterning, the codas and ambisyllabic situations were not reliably different (see also Gick, 2003b using ultrasound data). That is, the ambisyllabic context looked coda-like for these liquids and glides.

We wish to expand on these few earlier studies of stress effects on segments internally composed of multiple gestures to determine how stress affects segment-internal coordination. The early studies looked primarily at single gestures ([p], [t/ɾ]) or at [m] or liquids/glides; we will focus on alveolar nasals to determine whether these findings extend to gestural molecules composed of a tongue tip constriction gesture plus a secondary velum gesture. Further, we will extend the early findings by looking at additional word-internal stress environments and how they affect intergestural timing. We anticipate extending the understanding of word edge, specifically onset, effects for [n] from the first part of the study to word-internal position in the second part of the study.

Thus in summary, this study will utilize real-time vocal tract imaging to examine the coordination of velum and tongue tip for the production of [n] as a function of syllable position and stress. Specifically, we will investigate whether (1) two discrete *modes* of timing are observed within segments, reflecting coda and onset position, (2) if so, whether juncture geminates show one of these patterns or an intermediate pattern, (3) whether intervocalic word-internal [n]s have different internal coordination as a function of the stress pattern of their adjacent vowels.

2. Method

Real-time MRI was used to study the relative timing of the movements of the velum and tongue tip in the production of alveolar nasal consonants in three different syllable conditions at word edges, and, word-internally, in three different stress conditions.

2.1. Subjects

Four native speakers of American English participated in this experiment. Subjects will be referred to as Subjects A, J, K, and E; Subjects J (male) and A (female) are both from California, subject K (male) from Virginia, and subject E (male) from Georgia. All were healthy subjects over the age of 18, and none reported any speech, hearing, or language problems. All were paid for their participation in the study, which took about 1 h.

2.2. Stimuli

2.2.1. Stimulus set one—the syllable condition

The first stimulus set consisted of pairs of monosyllabic words containing an alveolar nasal in (i) syllable-initial position [oʊ#noʊ], (ii) syllable-final position [oʊn#oʊ], and (iii) intervocalically as a juncture geminate [oʊn#noʊ], spanning the word boundary. The phonological vocalic context surrounding the nasal was kept constant. Each stimulus was presented in two carrier sentences, one in which the target sequence was flanked by labial consonants on both sides and another in which the sequence was flanked by alveolar consonants, yielding 6 stimuli. The stimulus set and carrier sentences are shown in Table 1; we have explicitly manipulated position-in-word in order to exercise strong controls over position-in-syllable.

The stimuli were presented in block capitals to encourage subjects to utter both words with focal accents, thereby minimizing the chance of resyllabification of coda consonants as onsets. Subjects consistently produced glottal stops for the vowel-initial second syllables. Because of scanner noise, subject pronunciation was evaluated only post-hoc (see also Section 2.4). Nine onset tokens (labial frame) of one subject (E) were pronounced with the vowel [aʊ] rather than [oʊ]; it was decided to include these tokens in the analysis as there was no reason to think this vocalic pronunciation would be relevant to the dependent measures of interest; however, subject E's onset results can be evaluated with this pronunciation variant in mind. Regarding the relative prominence of the words, though both were written in caps and instructed so as to have equal (focal) accent, some difference in relative prominence is unavoidable (given that pausing was not occurring); the second word was somewhat more prominent for subjects A, E, and J, and any prominence difference was less clear for K with perhaps the first word being more prominent.

2.2.2. Stimulus set two—the stress condition

The second stimulus set consisted of single target words with word-internal nasals. The three stress conditions consisted of (i) an unstressed syllable preceding the target nasal and a stressed syllable following it, (ii) a stressed

Table 1
Syllable condition stimuli.

Syllable condition	Target sequence	Labial frame *I can type ____ five times*	Alveolar frame *I can write ____ six times*
Onset	[oʊ#**n**oʊ]	BOW **K**NOW	TOE **N**ODE
Coda	[oʊ**n**#oʊ]	BO**N**E OH	TO**N**E ODE
Juncture geminate	[oʊ**n**#**n**oʊ]	BO**N**E **K**NOW	TO**N**E **N**ODE

Author's personal copy

Table 2
Stress condition stimuli—note for cells in which there is a /, two subjects produced one form and two subjects the other.

Stress condition	Target sequence	Labial frame *I can type ____ five times*	Alveolar frame *I can write ____ six times*
Unstressed–stressed	[ə'**n**oʊ]	beg**n**ome/bek**n**ow	de**n**ote
Stressed–unstressed	['oʊ**n**ə]	bo**n**afide	to**n**ative
Primary stressed–secondary stressed	['oʊˌ**n**oʊ]	bo**n**o	no**n**o/o**n**o
From Table 1 above, repeated here			
Stressed–stressed (onset)	[oʊ#**n**oʊ]	BOW KNOW	TOE NODE

syllable preceding the target nasal and an unstressed syllable following it, and (iii) a syllable with primary stress preceding the target nasal and a syllable with secondary stress following it. (We have used "unstressed" to refer to a syllable pronounced with a reduced schwa vowel.) The stimuli are given in Table 2. All of these targets were presented in labial and alveolar segmental frame sentence contexts as in set 1, yielding 12 stimuli. (A complementary set of words with the same stress conditions but containing oral stop targets was additionally recorded but will not be analyzed here.)

For two cells of the six cells above, two subjects (J and A) produced a form with a word-final consonant (labial frame Unstr-Str; alveolar frame for Str-Unstr) in the target word, and two produced a vowel-final form of the target word (see cells in Table 2 with a slash). (Note that each subject is analyzed independently in the design). Each form has advantages and disadvantages: namely having the marginal consonant present is more like the other words in the experiment but also introduces another nasal in the vicinity. (We will see in fact that the subjects behaved similarly to one another in their timing pattern.)

Each subject, thus, read a total of 18 stimuli (6 of these are the oral stops not studied here). Ten repetitions of each stimulus were acquired in blocks. The stimuli in the stress set were further blocked by stress (3 groups), to avoid confusion for the subject. The order of the three stress groups and the syllable set was randomized for each repetition, as was the order of the stimulus items within each of these groups. All subjects read the same randomization.

2.3. Imaging

The MR images were acquired using fast gradient echo pulse sequences and a 13-interleaf spiral acquisition technique (Narayanan et al., 2004), developed within the RTHawk framework (Santos, Wright, & Pauly, 2004), on a conventional 1.5 Tesla scanner. The midsagittal slice thickness was 3 mm. The interleaf value of 13 means that 13 radio frequency excitation pulses were fired for the acquisition of a complete image. These excitation pulses were fired every 6.856 ms, resulting in a frame rate of 11 frames per second (fps), that is, one entire frame of new information every 89 ms (6.856 ms × 13). Reconstruction of the raw data was implemented using a standard gridding and sliding-window technique (Jackson, Meyer, Nishimura, & Macovski, 1991) with a window offset of 48 ms (the time that elapses between 7 successive excitation pulses). This is, that from the original data we have one frame for each 89 ms (13 acquisitions), but the frame rate is enhanced by combining data from the last 7 acquisitions of frame n, and the first 6 acquisitions of frame $n+1$. This produces a series of 68×68 pixel images, each of which contains information from the preceding frame and a proportion of new information, thus affording us with a frame rate of 21 fps (i.e., one image every 48 ms) for subsequent processing and analysis.[3] Data from two channels of a four-channel targeted phased-array receiver coil specifically designed for vocal tract imaging were combined by performing root sum of squares of the images. These two channels provided the desired information about the vocal tract areas of interest—the front of the face and neck. Available electronically with this paper are three sample real-time MRI movies of sentences spoken by subject J, one from each syllable condition. Further accounts of the general real-time MRI technique our group employs and challenges we've encountered can be found in: Narayanan et al. (2004), Bresch, Adams et al. (2006), Bresch, Neilsen et al. (2006), and Bresch, Kim, Nayak, Byrd, and Narayanan (2008).

2.4. Data collection

In this study, subjects were familiarized with the stimuli in order to prevent mispronunciations since some were non-words (e.g., *beknow, tonative*). They were instructed to speak at a normal rate and volume. The aforementioned coil, which was attached to a Perspex frame, was lowered so as to be close to the subject's face without touching it,

[3]Some new information on the vocal tract's entire geometry is obtained every TR ["repetition time": the time between two consecutive MR excitation pulses], that is, with a rate of roughly 160 Hz. However, the amount of information captured at every TR is not sufficient to produce a complete image, and hence a sliding window method is adopted. Successive images reconstructed with the sliding window technique at every TR would differ in max 8% (i.e., 1/13) of the information content. We chose a window offset of 6TR so that 46% of each image reflects new data, and the remainder is the same as from the previous image. Through this step we decrease the temporal resolution of our image-derived measurements with the benefit of a reduced edge tracking effort. Generally speaking, even short phenomena, i.e., about as long as one TR, will reflect in the images even with sliding window reconstruction.

Author's personal copy

and padding was placed around his/her head to prevent movement. Subjects lay prone in the scanner bore while producing the stimuli, and wore foam earplugs to protect the ears. Once inside the scanner, subjects attached an adhesive plastic folder containing the stimulus set, with each sentence and each page numbered, to the top of the bore. During the scan, subjects were prompted verbally with the number corresponding to each stimulus item over an intercom.

Each scan was limited to a duration of approximately 30–45 s, after which the scanner was allowed to cool for another 30 s. Most stimulus blocks were completed within one such period; however, it was rarely necessary to have subjects repeat the final sentence of one block as the first sentence in the subsequent scanning period or in a scanning period of its own. Fillers were not included to maintain the brevity of the stimulus blocks within the scanner interval.

In addition to the MRI data, simultaneous speech audio was also collected using a fiberoptic microphone.[4] A 10 MHz signal from the scanner's master clock was used to timelock the audio and video data acquisition. An additional scanner start-stop signal, which accompanies every 6.856 ms data acquisition interval, was utilized to ensure synchronicity of the audio and video sequences during subsequent reconstruction (Bresch, Neilsen et al., 2006). The audio signal was then low-pass filtered to 10 kHz and the sampling rate subsequently reduced from 100 to 20 kHz before being processed with the noise-cancellation procedure described in Bresch, Neilsen et al. (2006) to produce a relatively clear audio signal.

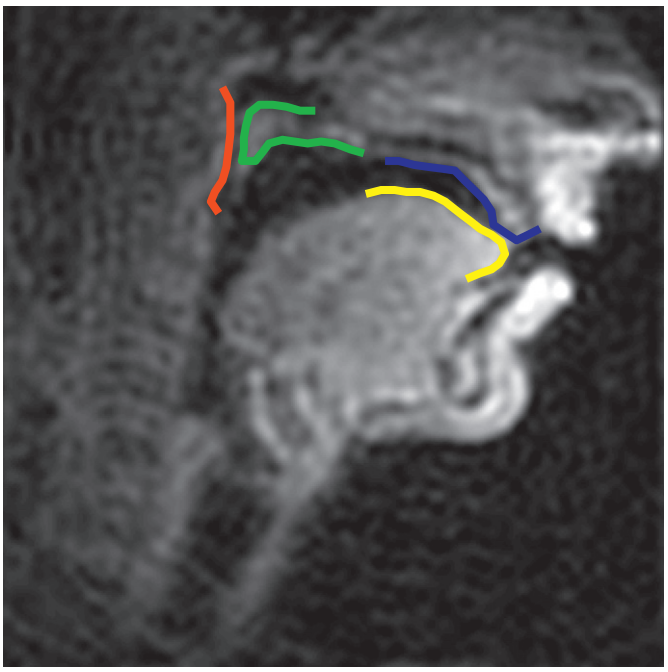

Fig. 1. A sample tracked frame from a data real-time MRI movie. The overlay lines mark the (time-varying) air–tissue interfaces in the midsagittal plane. The yellow line corresponds to the surface of the tongue body and tip, the blue line to the hard palate, the green line to the velum, and the red line corresponds to the section of the pharyngeal wall that can come into contact with the velum. In each image, the position of the individual air–tissue boundary segments was determined using a semi-automatic edge detection procedure (Bresch, Adams et al., 2006). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

2.5. Sequence selection, tracking, and aperture function calculation

The availability of synchronous audio facilitated the process of excising the appropriate intervals of the movie files, namely, the target sequence and the preceding and following words of the carrier sentence. An Active Contours algorithm (Bresch, Adams et al., 2006) was applied to the segmented MRI movies to track (i) the tongue, (ii) the alveolar ridge, (iii) the velum, and (iv) the pharynx wall.[5] In order to make the movie sequences compatible with the algorithm, each video image was upsampled by a factor of 5 using zero-padding in the Fourier domain. Points were initialized manually along the relevant contours of the first image of each sequence. An insertion/deletion algorithm then increased the number of points on the contour where resolution was too low. Fig. 1 shows a screen shot of a tracked frame from a data movie. Example real-time MRI movies of the data have been uploaded as Supplementary material.

For each frame of the movie, the minimum aperture between the tracked (i) tongue and alveolar ridge for the target tongue tip gestures, and between the tracked (ii) velum and pharynx wall for the target velum lowering gesture were calculated from the time-varying tracking

[4]The audio acquisition was carried out using a National Instruments DAQ 6036-E data acquisition card. While this 16-bit digital-to-analog converter card allows operating on an external sample clock, which is required for sample-exact synchronization and proper MRI noise cancellation, it lacks input anti-aliasing filters. Hence, a simple first-order analog anti-aliasing filter with a 3 dB cut-off at 9 kHz was added in the signal path and a high over-sampling rate of 100 kHz was chosen. Subsequent high-order low-pass filtering in the digital domain and down-sampling produce an output audio signal at a sampling rate of 20 kHz with a reasonably flat pass-band up to 9 kHz band width, which is used for all subsequent operations.

[5]The contours were tracked using the procedure described in Bresch, Adams et al. (2006). To summarize the contour tracking was carried out using the well-known SNAKE and Optical Flow algorithms. Starting with the contour location of the previous frame, the Optical Flow procedure is used to estimate the new position of the endpoints of the contour in the current frame. The SNAKE active contour algorithm is subsequently applied to find the optimum fit of the remainder of the contour to the current image. Hereby, the active contour method optimizes the contour's vertices' positions so as to align the contour with the maximum image intensity gradient while simultaneously attempting to mitigate the influence of image noise by penalizing the resulting curvature. The initial contour estimate for the first frame is entered by the user. In order to avoid the propagation of tracking errors, the initial guess for the contour location in any frame is checked and if necessary corrected by the user before the SNAKE algorithm is run.

Author's personal copy

contours. The data were then upsampled by a factor of 10 in the time domain[6] and low-pass filtered at 7 Hz (values chosen based on inspection of the time-varying characteristics of the aperture function). This allowed us to produce relatively smooth plots of aperture values over time, removing the risk of spurious fluctuations of aperture. Importantly, this entire process makes it possible to look at kinematic velum function non-invasively in real-time, which has not before been possible.

2.6. Data measurement, dependent variables, and statistical analysis

MVIEW (M. Tiede, in development) in the MATLAB computing environment was used to analyze the aperture functions. Time and aperture values were recorded at four points for both aperture functions—velum and tongue tip. First, the maximum (extremum) velum aperture and minimum (extremum) tongue tip aperture were identified based on zero-crossings in the aperture derivative. These are identified as the targets for the gestures. Secondly, a plateau around the target was defined as beginning and ending at the points at which the aperture values entered and left a value of 80% of the total gestural aperture range of that gesture. These two points define the target plateau for the gesture. Lastly, the extremum aperture value (minimum for velum and maximum for tongue tip, as defined by a zero-crossing in the aperture derivative) preceding the gesture target extremum was identified as the gesture onset. (In rare cases when the range between target and the nearest potential onset extremum was less than 50% of the total aperture range for the gesture, the next earlier extremum was identified as the gestural onset, so as to avoid labeling minor fluctuations in the aperture function incorrectly as the gestural onset.) These algorithmically marked points are shown in Fig. 2.

Two dependent variables of interest were calculated based on these time points in the aperture functions. The first is a dependent variable that captures the lag time between the velum gesture and the tongue tip gestures. **Lag** is the time between when the tongue tip gesture reaches (i.e., first enters) its target plateau region and the time when the velum gesture reaches (i.e., first enters) its target plateau region. This is calculated so as to be negative when the velum gesture precedes the tongue tip gesture; a nominal zero would indicate synchrony of the two targets, and a positive value would indicate that the tongue tip gesture reached its target before the velum gesture. The second dependent variable is VELDISP, which is the change in aperture—i.e., the displacement—from the onset of the velum gesture to its maximal aperture. This is an indicator of the spatial magnitude of the velum lowering gesture. These variables are shown in a single-token schema in Fig. 3.

On occasion other durational and spatial measures are referred to below to complement these two primary measures of interest.

A separate two-factor ANOVA is run for each subject for each stimulus set, with the factors of frame-sentence-context (bilabial/alveolar) and syllable or stress condition. All significant effects are reported. Significant ($p<.05$) main effects of the experimental variable (syllable or stress condition) are further investigated with a Scheffe post-hoc test ($p<.05$). In experiment 1—the syllable structure experiment—both the temporal and the spatial dependent variables are of interest in light of the experimental hypotheses. However, in the stress experiment, the particular interest is in how stress affects timing; so primarily the LAG variable will be examined, though significant patterns in the spatial domain will also be noted.

3. Results

3.1. Syllable position study

A two-factor ANOVA with the factors frame-sentence-context (alveolar, labial) and syllable condition (onset, coda, juncture geminate) was run for each subject on the timing variable LAG (recall that this is the timing lag between the plateau edge for velum lowering achievement and the plateau edge for tongue tip constriction achievement.) The frame-sentence-context variable was not of experimental interest; we report simply that one subject (A) had an effect of frame-sentence-context [longer negative lags in the alveolar frame $F(1,52) = 13.897$, $p<.001$]; there were no significant interaction effects.

All subjects had significant main effects of syllable condition on LAG ($p<.0001$; Subject A: $F(2,52) = 23.671$; E: $F(2,48) = 26.188$; J: $F(2,52) = 11.891$; K: $F(2, 52)$ 35.676). Post-hoc Scheffé's tests indicated that for Subject A, all three syllable conditions differed from one another, with coda and geminate having mean negative lags (i.e., onset of velum plateau before onset of tip plateau) of −62 and −25 ms, respectively, and the onset having a positive lag of 26 ms (standard error of 12 coda, 11 gem, 9 onset). The other three subjects all had significant post-hoc differences such that onsets differed from both codas and geminates in their lags, but codas and geminates did not differ: means and standard error in ms, for subject E: coda −90 (12), gem −88 (15), ons 4 (5); subject J: coda −57 (9), gem −63 (9), ons 2 (12); subject K: coda −86 (6), gem −90 (11), ons −1 (9). Because the speakers showed the same general pattern of results, a summary figure with speakers pooled is given in Fig. 4.

A nearly identical pattern of results obtains when the lag between onsets of the velum and tongue tip movements (ΔOnsets) is considered (though variability is considerably higher for this measure for every subject than for the LAG

[6]The upsampling was carried out by increasing the sampling rate by a factor of 10 through zero insertion and subsequent anti-alias (low-pass) filtering. Hereby, a linear-phase finite impulse response filter was used in order to avoid any dispersive/distorting effects on the time waveform.

Author's personal copy

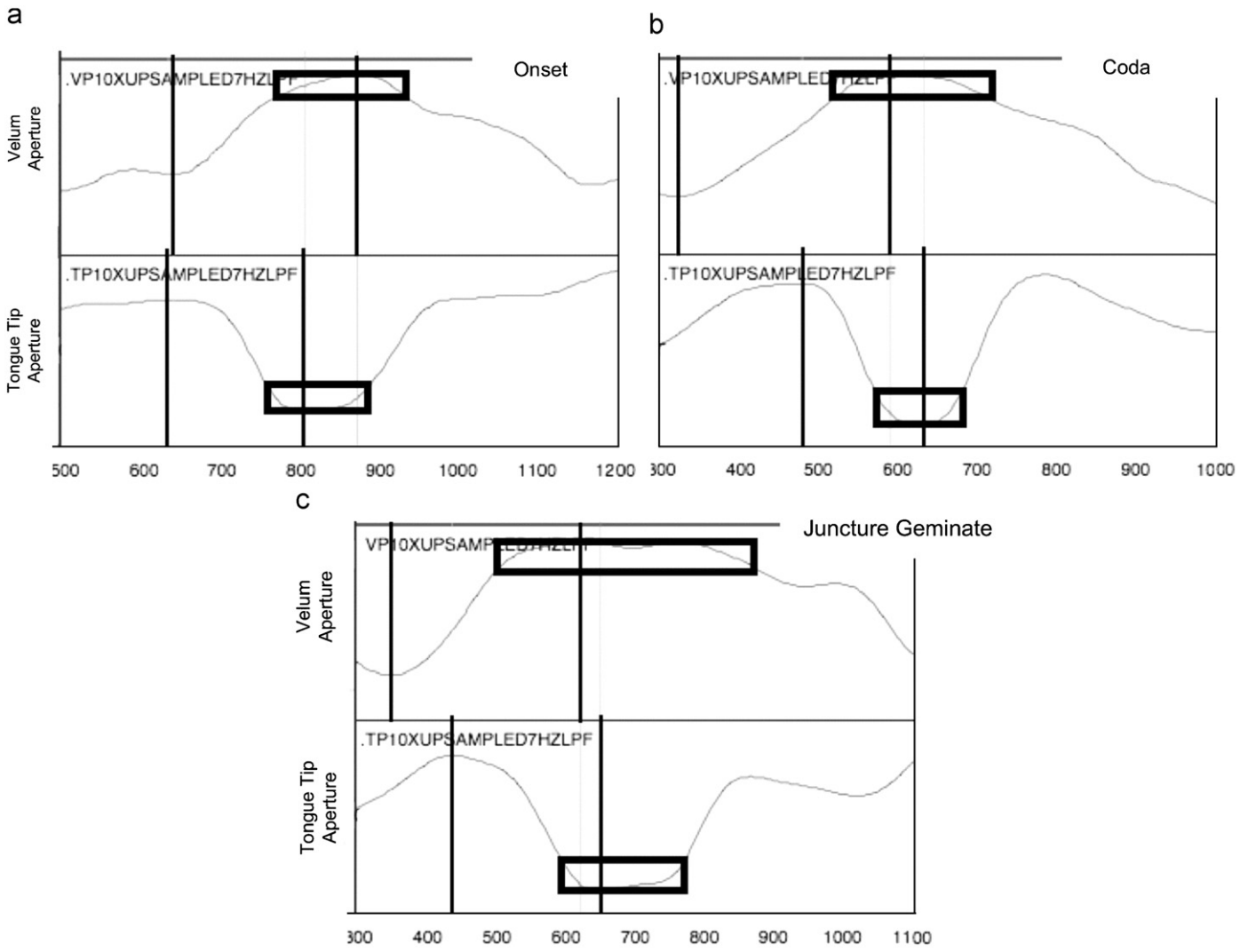

Fig. 2. Samples showing the marked time points for gesture onset (vertical line 1) and peak (vertical line 2) and the target plateau interval (rectangle). Velum aperture is in the top panel of each figure and tongue tip aperture in the bottom panel for each individual example.

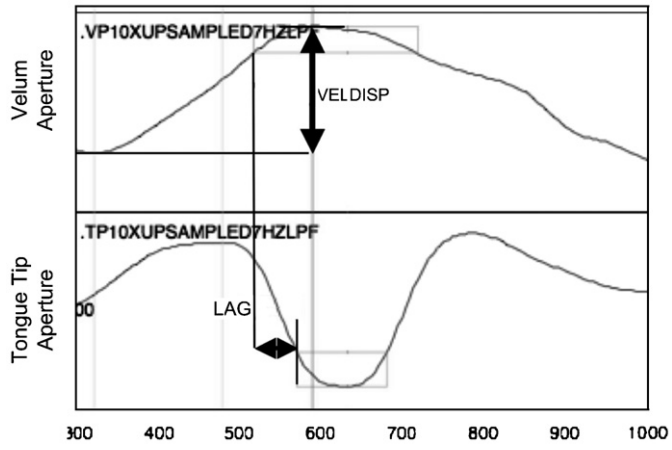

Fig. 3. A sample schematizing the dependent variables of velum aperture displacement (VELDISP) and velum–tongue tip lag (LAG). A negative LAG is shown in this example.

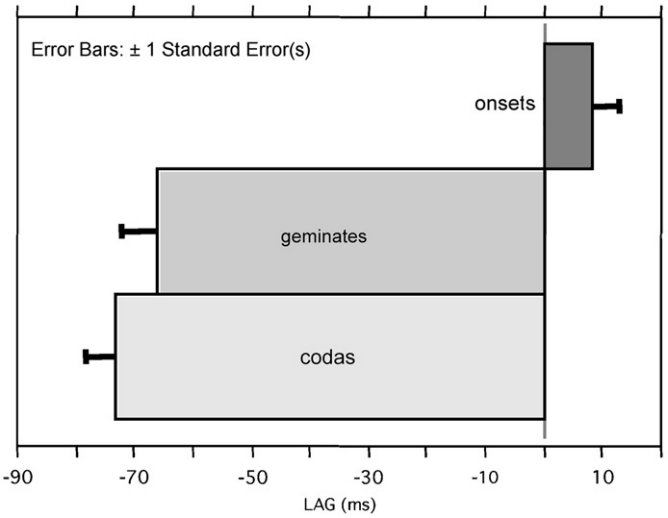

Fig. 4. Pooled results showing mean LAG for onsets, juncture geminates, and codas.

Author's personal copy

measure above): main effects: A: $F(2,52) = 19.816$, $p < .0001$ [frame-sentence-context $F(1,52) = 19.816$, $p < .0001$; alveolar frame longer negative lags]; E: $F(2,50) = 15.712$, $p < .0001$ [frame × syll interaction $p = .02$], J: $F(2, 52) = 10.286$, $p = .0002$; K: $F(2,52) = 35.922$, $p < .0001$ [frame-sentence-context $F(1,52) = 12.079$, $p = .001$; alveolar frame longer negative lags]. For all subjects the Scheffé's post-hoc tests show that onsets differ significantly from the codas and juncture geminates and that the latter two do not differ significantly: means and standard error in ms, for subject A: coda −113 (22), gem −72 (21), ons 48 (23); subject E: coda −158 (18), gem −137 (22), ons −33 (12); subject J: coda −149 (12), gem −145 (12), ons −32 (31); subject K: coda −162 (12), gem −155 (12), ons −27 (18). (The means are longer for this variable because constriction formation duration is longer for the velum than it is for the tongue tip.)

In sum, this result demonstrates the hypothesized bimodal timing pattern. Onsets show near synchronicity in the plateau achievements or, for one subject, a slight advance of the tongue tip constriction achievement relative to velum lowering achievement. Codas show a negative LAG in which velum lowering is achieved before tongue tip raising is achieved. This result is robust and consistent across subjects. Furthermore, the timing of velum lowering with respect to tip raising for the juncture geminates appears to be like that of codas. One subject showed a slightly shorter, but still negative LAG here, but in the majority, juncture geminate LAG were not different than that of codas. This is not surprising in that the lag measure used the constriction formation landmarks at the left edge of the juncture geminate, so would be expected to reflect, all else being equal, a coda-like LAG.

Next let us turn to VELDISP, which is an indicator of the spatial magnitude of velum aperture. Unlike other imaging, tracking, and transducing techniques like point-tracking techniques, in this case it is in fact specifically velum *aperture* from the pharyngeal wall that is being referenced in this study. One subject E had a significant effect of frame-sentence-context on VELDISP ($F(1,51) = 5.021$, $p = .0294$; larger VELDISP in the bilabial frame), but no subjects showed any interaction effect. Three subjects (A, J, K) had significant effects of syllable condition on VELDISP, and the fourth (E) had a marginal effect (A: $F(2,53) = 16.922$, $p < .0001$; E: $F(2, 51) = 2.937$, $p = .0621$; J: $F(2,52) = 12.239$, $p < .0001$; K: $F(2,52) = 21.915$, $p < .0001$). Speakers K and J both show a pattern in the post-hoc tests of onsets having smaller aperture displacement and geminates and codas not being distinct in (larger) aperture displacement. In contrast, Subject E shows no significant post-hoc differences, though her means do tend in this direction (post-hoc tests for onset differences fall at $p = .09$ and $p = .11$). Subject A shows smallest aperture displacement for onsets, intermediate displacement for codas, and large displacements for the juncture geminates; however, the post-hoc test between codas and onsets does not reach significance ($p = .09$); the other pairwise comparisons are significant.[7] In sum, the spatial magnitude differences as a function of syllable position are less systematic and robust than the timing differences, but there is still a clear effect of syllable structure on the velum aperture displacement. Onsets generally show the least change in velum aperture, as compared to coda and juncture geminate [n]s.[8]

In complement to this spatial measure of velum lowering, one might also want to consider the impact of the velum lowering gesture by examining how long the velum stays in its lowest posture as measured by the duration of the velum plateau. Whereas the codas and the juncture geminates did not show great differences in the degree of velum lowering, they do differ in the time spent at this lowest posture, with all subjects showing a main effect of syllable position ($p < .001$). Post-hoc tests show that all subjects except A have significantly longer velum plateaus for the juncture geminates than for onsets or codas. Subjects A and E also have a significant onset vs. coda difference such that onset velum plateaus are shorter than codas. (Means in ms, for subject A: coda 152, gem 168, ons 118; subject E: coda 204, gem 275, ons 146; subject J: coda 177, gem 266, ons 209; subject K: coda 154, gem 209, ons 140). Thus, while velum displacement (VELDISP) did not particularly differentiate codas vs. juncture geminates, velum plateau duration was, not surprisingly, longer for the juncture geminates.

3.2. Stress context study

For the word-internal intervocalic stimuli, we will present two analyses. First we will present two-factor ANOVAs separately for each subject with the factors frame-sentence-context (bilabial, alveolar) and stress condition. Stress conditions are coded as follows: 0 + 1 indicates unstressed followed by primary stress; 1 + 0 indicates primary stress followed by unstressed, and 1 + 2 indicates primary stress followed by secondary stress. In a second analysis the onset consonants from study 1 will also be included (coded 1 + 1) so as to compare if and how the word-internal [n]s, which are phonologically generally viewed as preferring to syllabify as onsets (since they are phonotactically valid onsets), differ from the word-edge onsets.

In the first analysis of LAG in the stress study, Subjects A and K have a main effect of sentence-frame-context [A: $F(1,36) = 7.517$, $p = .0095$; K: $F(1,46) = 5.46$, $p = .0239$;

[7] A nearly identical pattern of overall results obtains when the maximum velum aperture is considered rather than displacement.

[8] Lastly, for the sake of completeness, we report the tongue tip displacement data. All subjects showed an effect of segmental context, not surprisingly since one context was alveolar; no subject showed an interaction of context and syllable condition. Only subjects E and K showed an effect of syllable condition (E: $F(2,50) = 15.026$, $p < .0001$; K: $F(2,52) = 6.766$, $p = .0024$). Post-hoc tests show Subject E to have smaller displacements in codas than both other conditions and Subject K to have smaller displacements in coda than in juncture geminates.

Author's personal copy

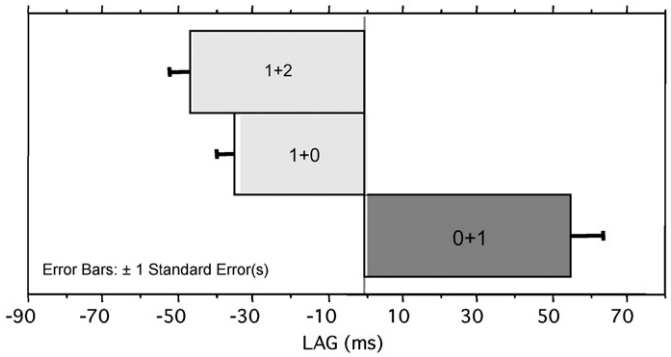

Fig. 5. Pooled results showing mean LAG for the three stress conditions: primary followed by secondary, primary followed by unstressed, and unstressed followed by stressed.

alveolar with larger negative lags for both], though Speaker E has an interaction effect [$F(2,37) = 17.504$, $p < .0001$, though the small number of available tokens in her alveolar 1 + 0 context make this impossible to interpret]. All subjects have significant main effects of stress condition: A: $F(2,36) = 10.326$, $p = .0003$; E: $F(2,37) = 32.518$, $p < .0001$; J: $F(2,38) = 18.374$, $p < .0001$; K: $F(2,46) = 24.467$, $p < .0001$. Further, all subjects show the same pattern of results in the post-hoc tests. For all, 0 + 1 differs from both 1 + 0 and 1 + 2, but 1 + 0 does not differ from 1 + 2. That is, weak–strong differs from strong–weak, but strong followed by unstressed does not differ from strong followed by secondary stress. An examination of the means shows that unstressed followed by stressed (0 + 1) always had a range of small to large positive lags (from 12 to 123 ms) [i.e., tongue tip plateau leading velum lowering plateau], while the conditions in which a primary stress syllable preceded the nasal had a range of small to large negative lags (from −7 to −70 ms). In sum, when the nasal is preceded by a stressed syllable, its coordination pattern is more coda-like,[9] but when it is followed by a primary (but not by a secondary stress), its timing pattern is more onset-like (Fig. 5).

Our second analysis will allow us to look at whether the weak–strong word-internal [n]s are discernibly different from the word-initial onsets by including the strong–strong word-onsets from study 1 as well. Since the 1 + 0 and the 1 + 2 were not significantly different above, they have been pooled here. This yields a three-level stress coding: weak–strong (0 + 1), strong–weak (1 + 0 and 1 + 2) and strong–strong (1 + 1, i.e., the word-onsets from Study 1).

The results of these additional two-factor ANOVAs on LAG indicate that three subjects (A, E, and K) have an effect of sentence-frame-context [A: $F(1,56) = 10.548$, $p < .002$; E: $F(1,52) = 14.418$, $p < .0004$; K: $F(1,66) = 15.684$, $p < .0002$; all with the alveolar context having LAGS more in the negative direction], and subject E again shows an interaction effect ($F(2,55) = 16.149$, $p < .0001$; primarily due to larger magnitude positive LAGS in the bilabial frame).

For the experimental main effect of stress condition, all subjects show a significant main effect: A: $F(2,56) = 20.379$, $p < .0001$; E: $F(2,55) = 31.993$, $p < .0001$; J: $F(2,56)$ 22.471, $p < .0001$; K: $F(2,66) = 29.511$, $p < .0001$. Post-hoc tests determine that for all subjects strong–weak is different from weak–strong in having negative vs. positive LAGS, which replicates the findings from the first analysis. For three of four subjects (A, J, K), strong–weak is also different from the strong–strong onset in this same way, and for the fourth subject (E) this effect also nearly reaches significance ($p = .0556$). Finally, for two subjects (E and A), weak–strong also differs from strong–strong in having larger magnitude positive LAGS. The mean LAGS (ms) are as follows: A: w–s 79, s–w −37, s–s 26; E: w–s 123, s–w −22, s–s 4; J: w–s 33, s–w −61, s–s 2; K: w–s 12, s–w −50, s–s −1.

In sum, the mean lags in the weak–strong environment were positive (indicating a later velum lowering relative to tongue raising), though K's only slightly so. In the strong–weak environment the mean lags were negative, and in the strong–strong environment they were near zero for three of the four subjects (indicating near simultaneous target achievement), and slightly positive for the other. However, only for two of the four subjects were the positive lags significantly different from the near zero lags found in the strong–strong environment.

Lastly, though we have no predictions regarding velum aperture as a function of stress, we include the stress effects on this dependent variable as a matter of empirical interest. Three subjects (A, E, and K) had a significant effect of the strong/weak stress environment on VELDISP. However, the pattern of their post-hoc tests was idiosyncratic, with A having strong–strong larger than the others in the alveolar frame, E having only weak–strong smaller than strong–weak, and K having strong–weak larger than the other two conditions. In sum, no consistent pattern for velum displacement emerges in the spatial domain as a function of stress.

4. Discussion

The study results above demonstrate that real-time vocal tract imaging using MRI can contribute to linguistic experimentation in a way complementary to other articulatory instrumentation. The findings help characterize the coordination of velum and tongue tip gestures for the production of [n] as a function of syllable position and stress. Specifically, we set out to investigate whether (1) two discrete *modes* of timing are observed to reflect coda and onset position, (2) whether juncture geminates show one of these patterns or an intermediate pattern, and (3) whether intervocalic word-internal [n]s have their internal coordination altered as a function of the stress pattern of their adjacent vowels.

[9] When LAGS codas from Study 1 are compared to 1 + 0 sequences from Study 2, two subjects show no difference in the negative lags; the other two (E and K) show that while both have negative LAGS, those of the word-final codas were larger in magnitude.

Author's personal copy

First, subjects consistently show two patterns of velum–tip coordination differentiating onset and coda nasals. In line with the empirical findings of Krakow (1989, 1993) and the theoretical proposal of Goldstein and colleagues (Goldstein et al., 2006; Goldstein et al., 2008; Nam et al., in press) the onset coordination appears to be one of synchrony or *in-phase* timing in which the gestures are triggered in rough temporal synchrony. The coda timing observed in the large negative lags suggests that, unlike in onsets, the gestures for the codas are triggered sequentially. In fact, the time from velum maximum aperture to the tongue tip gestural *onset* is short for the codas, much shorter than for the syllable-onset nasals in which the tongue tip onset is far earlier with respect to the velum extremum: Subject A: 50 ms vs. 178 ms; E: 57 ms vs. 126 ms; J: 57 ms vs. 160 ms; K: 25 ms vs. 125 ms.

In terms of the underlying coordination relations, it is difficult to determine without further modeling (and/or empirical data on stability) whether in-phase coordination is underlying specified for the vowel, tip, and velum in all pairwise combinations or whether, for example, both consonant gestures are individually coordinated (in-phase) to the vowel or whether the tip and velum are coordinated with one another and only the tip with the vowel. Since, barring other influences, these would all result in in-phase productions of the three gestures, they are all viable possibilities. Fig. 6 indicates all these possibilities along the top row, using the convention of solid line for in-phase timing. (Note that the vowel gesture is understood to be longer than the other gesture but is shown in the coupling graph with the same size icon for visual ease.) We reject for the time being the middle possibility in which the two gestures of the [n] are not coordinated with one another. We do this for the reasons outlined in Byrd (1996b)—namely, that it seems to be a well-founded assumption that the gestures composing a segment are coordinated with one another, in fact, that they are quite stably coordinated. (Consider how one would produce a click or properly contrastive VOT, for example, without a language-specific grammatical coordination of the component gestures [see e.g., Keating, 1984; Kingston & Diehl, 1994].)

The possible coda coupling relationships parallel these, but in this case antiphase coupling is presumed to exist between the preceding vocalic nucleus and its coda consonant (again we reject the logically possible middle scenario for the reason given above). (Recall that an antiphase coupling is when one action is triggered 180° out of phase with another action, yielding a sequential initiation of the actions due to the fact that the oscillators controlling their planning are coordinated such that one is initiating its cycle when the other is halfway through its cycle.) However, when that coda consonant is a multi-gesture consonant, as it is in this scenario, some questions arise: namely, are both gestures coordinated to the vowel or is only the more consonantal tongue tip gesture (see e.g., Gick, 1999)? We cannot resolve this here, but we tentatively lean toward the rightmost scenario in Fig. 5 having only the tip closure gesture directly coupled with its nuclear vowel. This coupling arrangement would seem to allow for more variability in amount of syllable-final vowel nasalization, since it lacks an instance of the more stable in-phase relationship that the leftmost scenario has and it also has fewer coupling relations in place, which increases variability (Nam et al., in press).

Goldstein et al. (2008) suggest that the coupling hypothesis provides an embodied theory of syllable structure grounded in fundamental characteristics of skilled human actions. They go on to elaborate how this

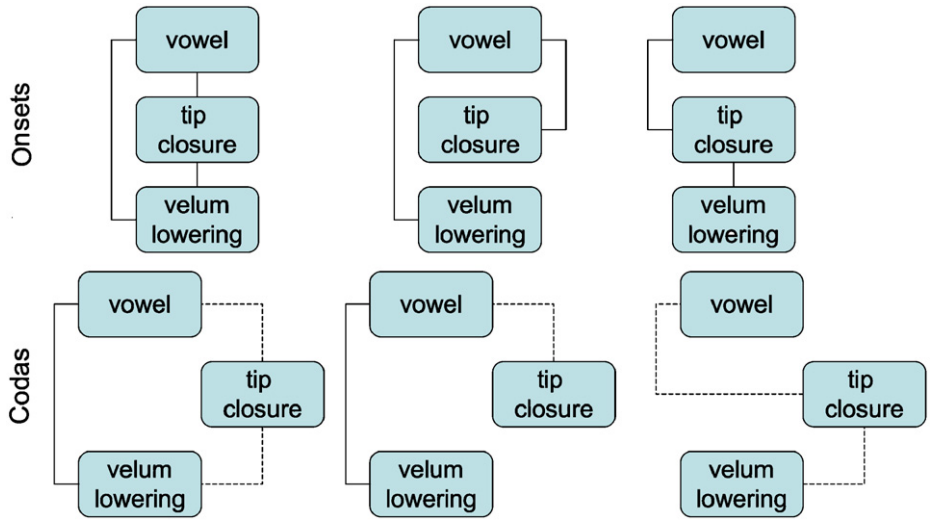

Fig. 6. Six possible coupling graphs: A schema indicating possible coupling relations in a syllable-onset nasal + vowel condition (top row) and a vowel + coda nasal condition (bottom row). The convention of solid lines for in-phase timing and dashed lines for antiphase timing (i.e., coupling) is used. (Note that the vowel gesture is understood to be longer than the other gesture but is shown in the coupling graph with the same size bar icon for visual ease. Also note the consonant–vowel timing was not examined in the empirical portion of this study.)

Author's personal copy

approach can explain cross-linguistic facts regarding: the universality of CV syllables (because the in-phase mode is more accessible and stable), increased C–C timing stability in onset clusters compared to coda clusters (Byrd, 1996a, b), the C-center effect (Byrd, 1995b; Honorof & Browman, 1995), the relatively freer combination of CVs than VCs, and the later emergence of onsets in phonological development. They suggest and model, however, that languages or structures within a language may differ on which gestures stand in the pairwise (in-phase or antiphase) gestural coupling arrangements. In particular, they suggest that both constriction (closure) gestures and release gestures might be possible topological nodes in the coupling graph of structures cross-linguistically such that languages have access to coordinating releases as well as closures and do so for phonologically significant ends (Goldstein et al., 2008).

Regarding our second question, which related to the segment internal timing pattern of juncture geminates, it seems clear from the results that the relationship between velum and tip target achievement are of the coda sort. This is sensible in that the "first" or "leftmost" consonant in the juncture geminate is of course a coda. Finally, only one subject showed a larger velum aperture displacement for the juncture geminates; most subjects showed no systematic difference from coda velum apertures, though juncture geminate velum plateaus were longer, as expected in the blended gesture.

Lastly, our third question related to whether the internal coordination of the [n] gestures is affected by the stress pattern of their adjacent vowels when the [n] is word-internal and intervocalic. The results clarify that the lexically stressed or contextually strong nucleus "attracts" the velum lowering gesture. The strong–weak context for the nasals behaves in a more coda-like fashion, consistent with other articulatory findings for consonants in such environments. While this and other articulatory experiments have focused on the strong–strong context or word-edge nasals in examining velum–oral coordination, it is clear that the finding of near simultaneous target achievement for these onset nasals must be tempered for intervocalic nasals so as to allow flexibility in the timing of velum lowering as a function of the local stress environment, as we see positive lags for the weak–strong intervocalic environment. Though we have not examined consonant–vowel timing here, this finding could be taken as tentative support for the top right phasing arrangement in the speculative Fig. 5, in which an in-phase (and therefore stable) timing relationship is lacking in the coupling graph *directly* between velum and vowel.

In sum, it is clear that syllable structure and stress interact to yield observed intergestural coordination patterns. Computational modeling work that incorporates a mechanism, or mechanisms, for capturing such interactions will be a critical next step in understanding the cognitive bases of speech timing. This has begun to be undertaken in work examining ensembles of coupled oscillators specifying phrase-level and foot-level timing and syllable- and foot-level timing (e.g., Barbosa, 2007; O'Dell & Nieminen, 1999; Saltzman, Nam, Krivokapic, & Goldstein, 2008). Saltzman et al. (2008) pursue a model of bidirectionally coupled phrase and foot planning oscillators and syllable and foot planning oscillators to capture the nested nature of the rhythmic system. Based on the empirical data they are able to model, they advocate for understanding an utterance's central clock as the result of a mutually entrained oscillatory ensemble (Saltzman et al., 2008). They also add an explicit influence of prosodic stress via a prosodic stress gesture or μ-gesture, modeled on Byrd and Saltzman's π-gesture (Byrd & Saltzman, 2003) for phrase boundaries that induces a local temporal modulation—a slowing—during its activation causing a stressed syllable to be longer than an unstressed syllable. It's not implausible (see e.g., Byrd & Saltzman, 2003) that this could yield intergestural overlap differences in the surface realization of the gestures. Further work on the integration of coupled oscillator models of speech timing with local temporal modulations required to correctly capture the durational properties of stressed syllables should help address how the timing among gestures is a combined result of syllable-level intergestural coupling with local prosodically induced lengthenings that instantiate stress.

5. Conclusion

In sum, we have demonstrated that real-time MRI can viably be used for reconstructing vocal tract apertures, allowing for the examination of parts of the vocal tract otherwise inaccessible to direct imaging or movement tracking. Using this technique to directly track velum and lingual apertures, we have been able to extend the seminal work of Krakow on intrasegmental timing English [m] to a consideration [n] in variable syllable positions and as a juncture geminate. We have found a bimodal timing pattern in which near-synchrony characterizes the timing for onsets and sequentiality the timing for codas, as described by Goldstein and colleagues (Goldstein et al., 2006; Goldstein et al., 2008; Nam et al., in press). Intervocalic word-internal nasals were found to have timing patterns that are sensitive to the local stress context, conforming to Krakow's suggestion that stress attracts the velum gesture. This suggests the presence of an underlying timing mechanism that can be influenced by or interact with the gestural specification of stress. We anticipate the need for further modeling work that will integrate accessible in-phase and antiphase coupling modes of gestural coordination for syllable structure, grounded in the characteristics of skilled human actions, with a timing model capturing stress and phrasal structure. Such a model has promise for capturing a range of empirical findings and illuminating our theoretical conception of the underlying timing principles at work in speech production.

Author's personal copy

Acknowledgments

The authors grateful acknowledge the support of NIH DC71243, the USC Imaging Science Center, and the assistance of Krishna Nayak, Jon Nielsen, and Daylen Riggs. The authors thank two anonymous reviewers and James M. Scobbie for their helpful comments.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.wocn.2008.10.002.

References

Barbosa, P. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, *49*, 725–742.

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Cambridge: Blackwell.

Bresch, E., Adams, J., Pouzet, A., Lee, S., Byrd, D., & Narayanan, S. (2006). Semi-automatic processing of real-time MR image sequences for speech production studies. In *Proceedings of the 7th international seminar on speech production*, Ubatuba, Brazil.

Bresch, E., Kim, Y.-C., Nayak, K., Byrd, D., & Narayanan, S. (2008). Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging. *IEEE Signal Processing Magazine*.

Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S. (2006). Synchronized and noise-robust audio recordings during realtime MRI scans. *Journal of the Acoustical Society of America*, *120*, 1791–1794.

Browman, C. P., & Goldstein, L. G. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Browman, C. P., & Goldstein, L. G. (1995a). Gestural syllable position effects in American English. In F. Bell-Berti, & L. J. Raphael (Eds.), *Producing speech: Contemporary issues (for Katherine Safford Harris)* (pp. 19–33). Woodbury, NY: AIP Press.

Browman, C. P., & Goldstein, L. G. (1995b). Dynamics and articulatory phonology. In R. F. Port, & T. Van Gelder (Eds.), *Mind in motion: Explorations in the dynamics of cognition* (pp. 175–193). Cambridge, MA: The MIT Press.

Browman, C. P., & Goldstein, L. G. (2000). Competing constraints on intergestural coordination and the self-organization of phonological structures. *Bulletin de la Communication Parlée*, *5*, 25–34.

Byrd, D. (1995a). Articulatory characteristics of single and blended lingual gestures. In K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIIIth international congress of phonetic sciences* (pp. 438–441).

Byrd, D. (1995b). C-centers revisited. *Phonetica*, *52*, 285–306.

Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*, 263–282.

Byrd, D. (1996b). A phase window framework for articulatory timing. *Phonology*, *13*, 139–169.

Byrd, D., Campos-Astorkiza, R., & Shepherd, M. (2006). Gestural de-aggregation via prosodic structure. In *Proceedings of the 7th international seminar on speech production*, Ubatuba, Brazil.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 139–169.

Clumeck, H. (1976). Patterns of soft palate movements. *Journal of Phonetics*, *4*, 337–351.

Cohn, A. (1990). *Phonetic and phonological rules of nasalization*. UCLA working papers in linguistics, 76.

de Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*, 283–310.

Delattre, P. (1971). Consonant gemination in four languages: An acoustic, perceptual, and radiographic study, Part I. *IRAL*, *IX/2*, 31–52.

Fougeron, C., & Keating, P. (1995). Demarcating prosodic groups with articulation. *Journal of the Acoustical Society of America*, *97*, 3384.

Fromkin, V. (1965). *Some phonetic specifications of linguistic units: An electromyographic investigation*. Ph.D. Dissertation, UCLA. UCLA Working Papers in Phonetics, 3.

Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech*, *33*, 195–258.

Fujimura, O., & Lovins, J. (1978). Syllables as concatenative phonetic units. In A. Bell, & J. B. Hooper (Eds.), *Syllables and segments* (pp. 107–120). Amsterdam: North Holland Publishing Company.

Fukaya, T., & Byrd, D. (2003). An articulatory examination of word-final flapping at phrase-edges and interiors. *Journal of the International Phonetic Association*, *35*, 45–58.

Gick, B. (1999). The organization of segment-internal gestures. In *Proceedings of the XIVth international congress of phonetic sciences*, San Francisco, August 1999.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI: Constraints on phonetic interpretation* (pp. 222–236). Cambridge: Cambridge University Press.

Gick, B., & Goldstein, L. (2002). Relative timing in the three gestures of North American English /r/. *Journal of the Acoustical Society of America*, *115*, 2481.

Giles, S. B., & Moll, K. L. (1975). Cinefluerographic study of selected allophones of English /l/. *Phonetica*, *31*, 206–227.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action in units understanding the evolution of phonology. In M. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2008). Coupled oscillator planning model of speech timing and syllable structure. In *Proceedings of the 8th phonetics conference of China and the international symposium on phonetic frontiers.*

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.

Honorof, D. N., & Browman, C. P. (1995). The center or edge: How are consonant clusters organized with respect to the vowel? In K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIIIth international congress of phonetic sciences*, Vol. 3 (pp. 552–555). Stockholm, Sweden: Congress Organisers at KTH and Stockholm University.

Horiguchi, S., & Bell-Berti, F. (1987). Aspiration, tenseness, and syllabification in English. *Language*, *47*, 133–140.

Huffman, M. (1997). Phonetic variation in intervocalic onset /l/'s in English. *Journal of Phonetics*, *25*, 115–141.

Jackson, J., Meyer, C., Nishimura, D., & Macovski, A. (1991). Selection of a convolution function for Fourier inversion using gridding. *IEEE Transactions on Medical Imaging*, *10*(3), 473–478.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Bloomington, IN: Indiana University Linguistics Club.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286–319.

Keating, P. (1995). Effects of prosodic position on /t,d/ tongue/palate contact. *Proceedings of the 13th international congress of phonetic sciences, Stockholm*, *3*, 432–435.

Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Kiritani, S., Hirose, H., & Sawashima, M. (1980). Simultaneous X-ray microbeam and EMG study of velum movement for Japanese nasal sounds. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, *14*, 91–100.

Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Doctoral Dissertation, Yale University, New Haven, CT.

Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate.

Author's personal copy

In M. A. Huffman, & R. A. Krakow (Eds.), *Nasals, nasalization and the velum (phonetics and phonology V)* (pp. 87–116). New York: Academic Press.

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

Munhall, K., & Lofqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, *20*, 111–126.

Nam, H., Goldstein, L., & Saltzman, E. (in press). Self-organization of syllable structure: A coupled oscillator model. In I. Chitoran, C. Coupe, E. Marsico, & F. Pellegrino (Eds.), *Approaches to phonological complexity*.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America*, *115*, 1771–1776.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. Docherty, & D. R. Ladd (Eds.), *Laboratory phonology*, Vol. 2 (pp. 261–280). Cambridge: Cambridge University Press.

O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences* (Vol. 2, pp. 1075–1078). Berkeley: University of California.

Ohala, J. J. (1971). Monitoring soft palate movements during speech. *Project in Linguistic Analysis*, *13*, J01–J015.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*, 499–526.

Saltzman, E., Lofqvist, A., & Mitra, S. (2000). "Glue" and "clocks": Intergestual cohesion and global timing. In M. B. Broe, & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 88–101). Cambridge: Cambridge University Press.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333–382.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the speech prosody 2008 conference*, Campinas, Brazil.

Santos, J. M., Wright, G. A., & Pauly, J. M. (2004). Flexible real-time magnetic resonance imaging framework. In *Proceedings of the 26th annual international conference of the IEEE EMBS*, September 1–5, 2004, San Francisco, CA, USA (pp. 1048–1051).

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*, 291–311.

Stone, M., & Hamlet, S. (1982). Variations in jaw and tongue gestures observed during the production of unstressed /d/'s and flaps. *Journal of Phonetics*, *10*, 401–415.

Turk, A. E. (1993). *Effects of position-in-syllable and stress on consonant articulation*. Ph.D. Dissertation, Cornell University.

Turk, A. (1994). Articulator phonetic cues to syllable affiliation: Gestural characteristics of bilabial stops. In P. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 107–135). Cambridge: Cambridge University Press.

Vassière, J. (1988). Prediction of velum movement from phonological specifications. *Phonetica*, *45*, 122–139.