Associations between tongue movement pattern consistency and formant movement pattern consistency in response to speech behavioral modifications^{a)}

Antje S. Mefferd^{b)}

Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, 1215 21st Avenue South, 8310 Medical Center East, Nashville, Tennessee 37232-8718, USA

(Received 22 January 2016; revised 18 August 2016; accepted 28 October 2016; published online 16 November 2016)

The degree of speech movement pattern consistency can provide information about speech motor control. Although tongue motor control is particularly important because of the tongue's primary contribution to the speech acoustic signal, capturing tongue movements during speech remains difficult and costly. This study sought to determine if formant movements could be used to estimate tongue movement pattern consistency indirectly. Two age groups (seven young adults and seven older adults) and six speech conditions (typical, slow, loud, clear, fast, bite block speech) were selected to elicit an age- and task-dependent performance range in tongue movement pattern consistency. Kinematic and acoustic spatiotemporal indexes (STI) were calculated based on sentence-length tongue movement and formant movement signals, respectively. Kinematic and acoustic STI values showed strong associations across talkers and moderate to strong associations for each talker across speech tasks; although, in cases where task-related tongue motor performance changes were relatively small, the acoustic STI values were poorly associated with kinematic STI values. These findings suggest that, depending on the sensitivity needs, formant movement pattern consistency could be used in lieu of direct kinematic analysis to indirectly examine speech motor control.

© 2016 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4967446>]

[JFL]

Pages: 3728–3737

I. INTRODUCTION

Although talkers rarely produce consecutive repetitions of exactly the same utterance in every-day life, such a speech task can be useful to examine the ability to reproduce a speech pattern generated for a specific utterance. So far, this ability has been studied mostly with motion capture technology to examine the speech movement pattern consistency of the lower lip and jaw. Based on these studies it is known that healthy mature talkers are able to produce highly consistent movements, whereas young children, aging adults, or impaired talkers produce comparatively inconsistent movement (e.g., Kleinow *et al.*, 2001; Green *et al.*, 2002; McHenry, 2003; Sadagopan and Smith, 2008; Grigos, 2009; Murdoch *et al.*, 2012; Schötz *et al.*, 2013; Grigos *et al.*, 2015). However, extremely consistent movements have also been observed in some talkers with speech motor impairments (Mefferd *et al.*, 2014), suggesting that impairment-related changes in movement pattern consistency can deviate from a healthy level in both directions.

Speech behavioral modifications are also known to impact speech movement pattern consistency. For example, in healthy talkers a slower-than-normal speech rate is commonly associated with low movement pattern consistency (e.g., Smith *et al.*, 1995; Kleinow *et al.*, 2001; Mefferd *et al.*, 2014). Taxing the motor system by asking healthy talkers to

speak as fast as possible can also result in low movement pattern consistency (e.g., Smith *et al.*, 1995; Kleinow *et al.*, 2001). Loud speech, by contrast, typically results in more consistent speech movements relative to habitual speech in healthy talkers (e.g., Dromey and Ramig, 1998; Mefferd and Green, 2010). Finally, practicing a novel speech task has been shown to result in increased movement pattern consistency (e.g., Walsh *et al.*, 2006; Sasisekaran *et al.*, 2010).

However, the interpretation of high and low movement pattern consistency remains difficult and depends on the theoretical perspectives of the examiner (Newell and Corcos, 1993). For example, some researchers interpret inconsistency as a source of error in motor planning and/or execution (e.g., motor schema theory, Schmidt, 2003; Schmidt and Lee, 2005). Others view inconsistency as a critical element of normal performance and regard only extreme levels of consistency and inconsistency as indicators of pathological processes (e.g., dynamic systems theory, Hausdorff *et al.*, 2000; Hausdorff, 2007; Stergiou and Decker, 2011). More to the point, it has been suggested that extremely high speech movement consistency may indicate rigid articulatory behavior in clinical populations such as amyotrophic lateral sclerosis (ALS; Mefferd *et al.*, 2014). By contrast, extremely low speech movement consistency has been associated with impaired execution, instability of the command sequence at the premotor stage, and numerous online adjustments (sub-movements) to the trajectory (e.g., Smith *et al.*, 1995; Wohlert and Smith, 1998; Kleinow *et al.*, 2001).

Regardless of the interpretation discrepancies, the examination of speech movement pattern consistency yields

^{a)}Portions of this work were presented at the Motor Speech Conference in Sarasota, FL, USA, in March 2014.

^{b)}Electronic mail: antje.mefferd@vanderbilt.edu

valuable information about the motor system and underlying motor control mechanisms. However, particularly tongue motor control, which is highly critical for speech production, remains understudied because expensive motion capture technology (e.g., electromagnetic articulography) is needed to capture articulatory movements. Further, most motion capture technologies for tongue movement tracking require the attachments of small sensors on the surface of the tongue. Due to this procedure, it can be difficult to study tongue movements in very small children, individuals with abnormal oral reflexes (e.g., in those with traumatic brain injury), and excessive involuntary head/body movements.

Because the acquisition of the speech acoustic signal is less invasive and less expensive than the acquisition of tongue kinematics, the speech acoustic signal is an attractive proxy for the study of tongue motor control. Recently, acoustic measures such as utterance-to-utterance spectrotemporal variation (UUV) and speech energy variation have been developed for such purposes (Howell *et al.*, 2009; Cummins *et al.*, 2014). Although these newly developed acoustic measures show promise as clinical tools, they combine phonatory and articulatory contributions, which makes it difficult to distinguish to what extent each subsystem contributes to the overall acoustic pattern consistency.

Articulatory movements, particularly those of the tongue, influence the formants in the speech acoustic signal (Fant, 1960). Thus, one approach to determine specifically articulatory control may be the assessment of formant movements. Particularly during vowel productions the first (F1) and second (F2) formants are directly associated with specific changes of vocal tract configurations. Movements of the tongue and jaw in the ventral-dorsal dimension, for example, are known to affect changes in F1, whereas tongue movements in the anterior-posterior dimension are associated with changes in F2. In addition, positional changes of the lips (e.g., protrusion), as well as laryngeal height adjustments and changes in pharyngeal shape can impact F1 and F2 values of vowels (Lindblom and Sundberg, 1971).

However, incremental changes of the vocal tract are known to show nonlinear formant acoustic changes (quantal relations; Stevens, 1972, 1989), which could weaken the predictability of speech movement pattern consistency based on formant movements. That is, small changes of tongue position can result in disproportional large changes in formants or can remain undetectable. Furthermore, talkers can change articulatory strategies while still achieving the same speech acoustic signal (motor equivalence). For example, talkers are known to trade tongue height and lip protrusion during the production of the vowel /u/ (Perkell *et al.*, 1993). It is, therefore, possible that the consistency of formant movements will be only poorly associated with that of tongue movements.

The strength of relation between tongue kinematic and formant acoustic movement consistency has been examined previously. Mefferd and Green, for example, have analyzed tongue movements and formant trajectories associated with the diphthong /ia/. Speech modifications (fast, loud, and slow speech) were used to elicit changes in tongue movement consistency and formant movement consistency. The

spatiotemporal index (STI; Smith *et al.*, 1995) was used to quantify movement consistency for tongue and formant movements. Although the kinematic STI has been frequently used in the past to quantify speech movement pattern consistency, a parallel formant acoustic STI measure had not been developed prior to the study by Mefferd and Green. Their findings showed that although speech modifications elicited variance in the tongue kinematic STI, task effects were relatively small and formant movement patterns did not reflect these task-related changes. Based on their findings the acoustic signal was thought to be insensitive to subtle changes in the kinematic signal.

However, the kinematic STI is typically calculated based on sentence-length movement signals to achieve an appropriate level of signal complexity (multiple peaks and troughs). Thus, weak mappings of kinematic and acoustic STI values in the study by Mefferd and Green (2010) may be due to the use of a very short and simple signal (the transition slope associated with the diphthong /ia/), which may have resulted in ceiling effects characterized by minimal task-induced changes in kinematic and acoustic STI values. More to the point, when sentence-length lip movement trajectories were used to calculate a lip kinematic STI, the mean STI values of young adult talkers ranged from approximately 11 to 18 across speech tasks (e.g., Kleinow *et al.*, 2001). By contrast, when diphthong transitions were used to calculate a tongue kinematic STI, the mean STI values ranged from approximately 5 to 7.5 across the same speech conditions (Mefferd and Green, 2010). The small changes in tongue kinematic movement consistency likely prevented a proper examination of articulatory-acoustic correlations. Therefore, a new study is warranted to re-evaluate the strength of articulatory-to-acoustic relation with a larger range in tongue kinematic STI values. The use of a sentence-length speech utterance, similar to the one commonly used in previous lip kinematic studies, may reduce the likelihood of ceiling effects in response to speech modifications because of the added sentence-length signal complexity compared to the simple diphthong transition used in the previous study by Mefferd and colleagues.

Therefore, the overall purpose of the current study was to examine the strength of association between tongue movement pattern consistency (quantified by the tongue kinematic STI) and formant movements (quantified by the formant acoustic STI) based on a sentence-length signal. Also in contrast to the previous study, which only studied young adult speakers, this study will include an additional group of very old speakers. The rationale for adding an older age group was based on findings that aging can negatively affect speech movement pattern consistency (Wohler and Smith, 1998; Kleinow *et al.*, 2001). Therefore, younger and older adults were included to create a broad performance range of tongue movement pattern consistency to test the strength of articulatory-to-acoustic relations across talkers at habitual speech.

As a second aim, this study sought to determine if known speech modification effects on tongue movement pattern consistency also elicit predictable changes in formant movement pattern consistency. Although it is known that in

healthy talkers tongue movement pattern consistency increases in response to slow or fast speech and decrease in response to loud speech (Smith *et al.*, 1995; Dromey and Ramig, 1998; Kleinow *et al.*, 2001; McHenry, 2003; Mefferd and Green, 2010), it remains unknown if such changes are also reflected in the consistency of formant movement patterns when more complex sentence-length signals are used. In addition to the speech conditions that have been previously included, clear speech will be added for the purpose of this study. Clear speech is also known to elicit change in speech movement pattern consistency. So far, clear speech effects on movement pattern consistency have not been studied in healthy talkers, however, in a study on talkers with Parkinson's disease decreased lip movement pattern consistency was reported relative to typical speech (Dromey, 2000). Although it is possible that clear speech effects may be larger in magnitude in talkers with Parkinson's disease than in healthy controls, a decrease in movement pattern consistency was also expected.

Finally, bite block speech will be included as an additional experimental condition. Presumably, the presence of a bite block perturbs the speech motor system and elicits a decrease in speech movement pattern consistency due to the novelty of the motor task (Schmidt and Wrisberg, 2004). That is, talkers rarely produced speech while holding a bite block in place. Particularly when participants are not allowed to practice prior to data collection, the first few sentence repetitions may be used to explore various articulatory adaptations to the bite block before selecting the most economical strategy that still achieves the desired speech acoustic signal (Lindblom, 1990). Based on previous studies, it is known that the presence of a bite block requires larger tongue displacements (or "supershapes" of the tongue) during high vowels and stops in order to overcome the unnaturally low position of the jaw and achieve perceptually adequate speech (Gay *et al.*, 1981). However, when comparing lip movement pattern consistency during typical speech (jaw free) and bite block speech, no significant changes were found (Namasivayam *et al.*, 2008). The researchers speculated that the relatively simple nature of the speech movement trajectory, which was based on bisyllabic pseudowords, may have resulted in ceiling effects. The increased signal complexity at sentence-length in the current study may prevent such ceiling effects. Change in tongue movement pattern consistency in response to bite block speech would allow us to test the strength of relations between changes in tongue kinematic STI values and those based on formant movement pattern.

II. METHOD

A. Participants

This study included 14 participants that were divided into two age groups. The young adult group comprised seven typical talkers (three females, four males) with a mean age of 23.3 yr and an age range from 18 to 26 yr. The older adult group consisted of seven typical talkers (three females, four males) with a mean age of 82.0 yr and an age range from 75 to 92 yr. All participants spoke English as their first and only language and did not report a history of a neurological

condition or a history of speech or language impairment. All participants in the older adult group passed the Mini Mental Status Examination (MMSE-2; Folstein and Folstein, 2012) with either 29 or 30 out of 30 points. All young adults passed a hearing screening at 0.5, 1, 2, and 4 kHz at 25 dB in both ears, five older adults passed a hearing screening at 0.5, 1, and 2 kHz at 35 dB in at least one ear, and two older adults wore bilateral hearing aids due to aging-related hearing loss. Given the primary aim of this research project it was not deemed necessary to exclude talkers with aided hearing. Further, an inspection of these individuals' data indicated that their performance did not differ from others in that group.

B. Experimental tasks

Participants were asked to repeat the sentence "Buy Kaia a kite" three times at their habitual rate and loudness to familiarize themselves with the target utterance. The utterance was designed to experimentally control for potential influences of lip rounding on formants, which would have been incompatible with the purpose of this study on tongue movement pattern consistency. After three practice trials, all participants completed a block of ten repetitions at the following speech conditions: typical, slow, loud, clear, fast, and bite block speech. The order of these speech conditions was randomized.

A magnitude production paradigm was used to elicit slow and loud speech (Tjaden and Wilding, 2004; Mefferd and Green, 2010). That is, participants were asked to produce speech that is approximately half their typical speaking rate by stretching their words rather than pausing between the words. For loud speech, participants were asked to speak approximately twice as loud as they normally do during typical speech. For fast speech, participants were asked to speak as fast as possible, whereas for clear speech participants were asked to repeat the target sentence as clearly as possible. To elicit clear speech, the experimenter described a situation of someone asking to the participant to speak more clearly because they could not understand what he/she said. For bite block speech, all participants were instructed to speak as they typically would while holding a bite block between their right molars. The bite block was created prior to data collection for each participant according to the procedures described by Netsell (1985). Briefly, participants were instructed to close their mouth while pressing the right molars into a small piece of condensation putty (Henry Schein, Melville, NY) until their front teeth reached an inter-incisal distance of 10 mm. After a short resting period, participants were asked to carefully open their mouths for the experimenter to remove the bite block and place it on a tissue to harden. To collect the speech movements for the bite block condition, the bite block was placed between the molars, proper positioning was checked, and participants were asked to produce the sentence repetitions without any opportunity to familiarize themselves with the bite block condition.

C. Kinematic data

Speech movements were captured using the three-dimensional (3D) electromagnetic articulograph (AG500,

Carstens Medizinelektronik GmbH, Bovenden, Germany). The AG500 has been shown to record orofacial movements with submillimeter spatial precision (Yunusova *et al.*, 2009). To capture tongue movements, two small sensors were attached with non-toxic dental tissue adhesive (PeriAcryl[®] 90, Glustitch, Delta, BC, Canada) to the surface of the tongue approximately 1 cm and 4 cm from the tip of the tongue at the sagittal midline. However, only the data of the posterior sensor were used for the purpose of this study. Based on previous investigations, the posterior tongue sensor moves the most during the vowel productions embedded in the target sentence. Further, to track head movements three or four reference sensors were placed on rigid goggles worn by the participant during data collection. Kinematic data were sampled at 200 Hz.

Using the *Normpos* software program (Carstens Medizinelektronik, Bovenden, Germany), 3D movements of the posterior tongue were corrected for head movements. The principal component signal was then calculated based on the three movement dimensions (see Fig. 1) and smoothed with a 15 Hz low pass filter using SMASH, a custom-written MATLAB program for the analysis of speech kinematics (Green *et al.*, 2013). The principal component signal of the entire sentence was used to determine tongue movement pattern variability. This approach has been used previously to reduce 2D or 3D kinematic data into one dimension (e.g., Adams *et al.*, 1993). As can be seen in Fig. 1, the onset and offset of each repetition was indicated by trough and peak in the principal component signal, which were associated with /a/ in “buy” and /i/ in “kite,” respectively.

The kinematic STI was calculated to quantify tongue movement pattern consistency (Smith *et al.*, 1995). To explain the STI calculation briefly, the ten principal

component signals associated with tongue movements of one speech condition produced by one talker were submitted to a customized MATLAB algorithm, which spatially and temporally normalized these signals. Then, 50 non-overlapping windows with 20 consecutive data points were created along the x axis and the standard deviation (SD) within each window was calculated. Finally, the sum of 50 SD values provided the STI score.

For the purpose of this study it was imperative that the kinematic and acoustic signal processing is paralleled in order to test articulatory-acoustic relations. Typically the STI is calculated based on the 3D Euclidean distance signal between a head reference sensor and the articulator (e.g., Smith *et al.*, 1995; Kleinow *et al.*, 2001; McHenry, 2003; Grigos, 2009; Nip *et al.*, 2009); however, a 3D Euclidean distance signal cannot be calculated for the speech acoustic data. Therefore, the principal component signal approach was chosen. The percent of data accounted for by the principal component signal was high across all participants and all speech tasks ($M = 89.6$, $SD = 8.15$). The suitability of this signal for STI calculation was further tested by submitting STI values derived from the principal component signals and those derived from the 3D Euclidean distance signals (calculated based on the center head sensor and the posterior tongue sensor) to a correlational analysis. A strong association was found [$r(85) = 0.98$] between the two signal approaches supported and validated the principal component approach used in this study.

D. Acoustic data

The acoustic signal of the speech utterances was recorded using a lavalier microphone (Audiotechnica AT899, Stow, OH) that was mounted 15 cm away from the mouth on the housing of the 3D electromagnetic articulograph. The audio signal was recorded at 16 kHz sampling rate, which was constrained setting by the manufacturer for the synchronized recording of acoustic and kinematic data. However, because only F1 and F2 formant values were of interest in this study, the sampling rate of the acoustic data is sufficient.

Using the spectrographic view in TF32 (Milenkovic, 2005) each repetition of the target sentence was parsed based on the first glottal impulse after the stop plosive /b/ in “buy” (onset) and the last glottal impulse before the stop plosive /t/ in “kite” (offset). The time histories of the first and second formants (F1 and F2) for all vocalic nuclei of the target section were obtained using the linear predictive coding (LPC) algorithm in TF32. LPC-based formant tracking errors were corrected manually. Formant values that were provided by the LPC algorithm were erased if they occurred outside of the vowel boundaries.

F1 and F2 time histories of the entire target utterance were extracted and submitted to a customized MATLAB algorithm where they were down-sampled to a sampling frequency of 100 Hz. Further, a piecewise cubic hermite interpolation function was applied to the F1 and F2 time histories to fill formant data during consonant productions and create a continuous signal (see Fig. 2). The interpolation within the sentence utterance was necessary to create a continuous movement

FIG. 1. (Color online) Time history of posterior tongue movement in medial-lateral (X), anterior-posterior (Y), and ventral-dorsal (Z) dimensions and the resulting principal component signal (PC).

pattern similar in complexity to the tongue kinematic signal. The interpolation should not interfere with the purpose of the study because when formants trajectories are highly consistent across repetitions, then the interpolation function also generates highly consistent interpolations for these repetitions.

The piecewise cubic Hermite interpolation (pchip) was selected because of its frequent use in geometric modeling of curves and motion trajectory (de Boor *et al.*, 1987). Its performance was first compared to other commonly used interpolation functions (e.g., cubic spline, full degree polynomial interpolation). Specifically, it was important that the interpolation between vowel-related formant trajectories would not create additional landmarks (peaks, troughs) in the signal.

Because of the inverse relation between tongue height and F1, F1 movements were inverted before a principal component signal was calculated based on F1 and F2 interpolated time histories (see Fig. 3). Similar to the kinematic principal component signal, the percent of data explained by the acoustic principal component signal was very high for all participants and all speech conditions ($M = 91.04$, $SD = 5.8$) and comparable to the percent of the kinematic principal component. The acoustic principal component signal was smoothed using a 15 Hz low pass filter. The acoustic STI was calculated using the same procedures that were implemented to calculate the kinematic STI.

Inter-rater reliability of the formant acoustic STI measure was determined based on the re-calculations of 50 of 84 acoustic STI values by a second examiner. Inter-rater reliability was found to be very high except in one case with a large discrepancy because of an error in tracking F3 instead of F2 for parts of the utterance. Thus, excluding this one data point from the inter-rater reliability evaluation, a bivariate correlation of measured and re-measured data resulted in

FIG. 2. First and second formant movements during the target utterance before interpolation (top) and after interpolation (bottom).

FIG. 3. Time history of the interpolated first and second formant movements (F1 inverted) and the corresponding principal component signal.

an inter-rater correlation of $r(49) = 0.92$; however, when including this data point, the correlation decreased to $r(50) = 0.84$.

Finally, to determine speaking rate changes, utterance durations were measured based on the tongue kinematic signal from the time of the movement onset to the movement offset as described above (see Sec. IIC). To determine loudness changes, vocal intensity was measured by extracting the mean intensity of each sentence repetition provided by the statistical output function in *Wavesurfer* (Sjölander and Beskow, 2011). A 440 Hz pure tone signal, which was recorded at 70 dB with the same amplification and the same distance from the loudspeaker as the mouth-to-microphone distance of the speech recordings, was used as a reference signal to calculate the sound pressure level (SPL) of the speech utterances. The intensity of the pure tone recording was verified with an SPL meter positioned directly next to the recording microphone.

E. Missing data

Although STI calculations were based on 10 repetitions for most participants, in 6 of the 84 calculations only 9 repetitions were used. Specifically, in 1 younger and 2 older adults, the initial repetition was excluded from STI calculations because it was produced in a different speaking style (more pausing, much slower rate) than all other repetitions. Further, in one younger adult and one older adult one repetition was excluded because of a recording error or because of a poor acoustic recording. Finally, in one older adult one repetition was excluded due to mispronunciation of a word in the target sentence.

III. RESULTS

A. The associative strength between tongue articulatory and formant pattern consistency

The top panel of Fig. 4 addresses the first aim of this study and shows the relation between the degree of tongue kinematic and formant acoustic movement pattern consistency at typical speech across younger and older adults in an

x/y scatter plot. A higher STI value indicated relatively low movement pattern consistency, whereas a lower STI value indicated relatively high movement pattern consistency. To determine if there was a systematic relationship between the observed variance in the tongue kinematic and formant acoustic STI values, all participants' data associated with typical speech were submitted to a Pearson's correlational analysis. Findings revealed a significant positive association between tongue kinematic and formant acoustic pattern consistency [$r(14) = 0.92$, $p < 0.001$] with 85% of the variance in formant movement pattern consistency accounted for by the variance in tongue movement pattern consistency.

The bottom panel of Fig. 4 addresses the second aim of the study and shows the relation between changes in tongue kinematic and formant acoustic movement pattern consistency in response to various speech conditions (i.e., clear, slow, loud, fast, typical, and bite block speech). Data points represent each participant's STI value at a given speech condition. All STI values were submitted to a Pearson correlational

analysis. Results yielded a significant positive association between tongue kinematic and formant acoustic movement pattern consistency [$r(84) = 0.64$, $p < 0.001$] with 41% of the variance in formant movement pattern consistency accounted for by the variance in tongue movement pattern consistency. However, this associative strength increases when just considering older adults [$r(42) = 0.74$, $p < 0.001$] accounting for 55% of the variance and it decreases when just considering younger adults [$r(42) = 0.60$, $p < 0.001$] accounting only for 34% of the variance.

Taking a closer look at each participant's data, the strength of association between tongue kinematic and formant acoustic STI values was examined by submitting the data to individual Pearson correlational analyses. The resulting correlation coefficients values are displayed in Table I. It can be seen that r -values varied greatly across talkers with moderate to strong associations in five younger and five older talkers and very weak associations in two younger and two older talkers (Cohen, 1988). However, it was also observed that for several talkers one speech condition greatly affected the overall relation. For example, the bite block condition had a considerable negative influence on the relational strength between articulatory and acoustic STI values for talker Y1, Y3, and O8. Similarly, the strength of this relation was tremendously improved for talker Y6 when the data of clear speech was not included in the correlational analysis (see Table I).

B. Speech task effects on kinematic and acoustic movement pattern consistency

Due to the small sample size nonparametric statistics were first considered. However, because results did not differ from those revealed by parametric testing, the parametric results are reported here. A repeated measures analysis of variance (ANOVA) revealed that the speech task had a significant main effect on tongue kinematic STI values [$F(5,60) = 5.788$, $p < 0.001$]; however, a main effect of age on tongue movement pattern consistency was not significant. Further, the age $\times$ task interaction was not significant. Figure 5 displays group means [$\pm$ SEs (standard errors)] for tongue movement pattern consistency (top) and formant movement pattern consistency (bottom) as a function of speech tasks. Brackets indicate significant pairwise comparisons. Specifically, significantly less consistent tongue movements were observed during slow speech relative to typical, fast, clear, and loud speech ($p \leq 0.04$). In addition, tongue movements were significantly less consistent during bite block speech than during loud and fast speech ($p < 0.01$). Finally, tongue movements were significantly less consistent during clear speech than during loud ($p = 0.01$) and fast speech ($p = 0.04$).

A repeated measures ANOVA revealed that the speech task had a significant main effect on formant acoustic STI values [$F(5,60) = 2.821$, $p < 0.024$]; however, a main effect of age on formant movement pattern consistency was not significant. Further, the age $\times$ task interaction was not significant. Pairwise *post hoc* comparisons yielded significant differences between loud speech and fast, slow, typical, and

FIG. 4. Association of kinematic and acoustic STI measures based on performance of younger and older adults during typical speech (top) and during various speech behavioral modifications (bottom).

TABLE I. Summary findings of individual's Pearson's correlation analyses including all speech conditions ($n = 6$) or excluding a speech condition that negatively affected the tongue kinematic-to-formant acoustic relation ($n = 5$). Also included is the range in change of the kinematic STI in response to speech condition. (** $p < 0.001$; * $p < 0.05$.)

Participant	Range of kinematic STI	Pearson's r $n = 6$	Pearson's r $n = 5$	Excluded condition
Y1	11.8	0.38	0.81	Bite block
Y2	16.98	0.98**	—	
Y3	13.96	0.56	0.97*	Bite block
Y4	4.67	0.15	—	
Y5	13.62	0.71	—	
Y6	9.71	-0.01	0.89*	Clear speech
Y8	8.54	0.55	—	
O2	11.91	0.94**	—	
O3	9.49	0.75*	—	
O4	6.44	0.54	—	
O5	10.89	0.81	—	
O6	6.02	0.88*	—	
O7	7.17	0.50	—	
O8	4.87	0.08	0.44	Bite block

bite block speech with formant movement patterns during loud speech being significantly more consistent than those associated with all other speech conditions ($p \leq 0.01$). Finally, formant movement patterns during slow speech were less consistent than those during clear speech ($p = 0.02$).

FIG. 5. Means ($\pm$ SE) of the kinematic STI (top) and the acoustic STI (bottom) across speech tasks.

C. Duration and vocal intensity changes in response to speech conditions

The top panel of Fig. 6 displays the mean utterance durations for each group and each speech condition. To determine between-group and within-group effects on utterance duration a repeated measure ANOVA was conducted. Further, Greenhouse-Geisser adjustments were applied because Mauchly's test of sphericity was significant. A main effect of speech task on utterance duration [$F(1.34, 16.11) = 70.68$, $p < 0.001$] and a significant task $\times$ group interaction [$F(1.34, 16.11) = 4.83$, $p < 0.034$] were found. The main effect of group was not significant. *Post hoc* pairwise comparisons yielded significant within-group effects on utterance duration in both groups. In the young adult group, fast speech was produced with significantly shorter utterance durations than all other speech conditions ($p \leq 0.007$), whereas slow speech yielded significantly longer utterance durations than all other speech conditions ($p \leq 0.001$). Further, clear speech was significantly slower than loud, typical, and bite block speech ($p \leq 0.009$), and loud speech was significantly slower than typical speech ($p \leq 0.02$). Older adults produced fast speech with significantly shorter utterance durations than all other speech conditions ($p \leq 0.02$), whereas slow speech yielded significantly longer utterance durations than all other speech conditions ($p \leq 0.004$). No significant differences in utterance duration were found for any other comparisons across speech conditions. With regard to between-group comparisons, younger adults demonstrated significantly shorter utterance durations than older adults during fast speech ($p = 0.02$). Further, utterance durations tended to be shorter in younger adults than in older adults during typical speech ($p = 0.05$).

Mean vocal intensities for each speech condition are displayed in the bottom panel of Fig. 6. A repeated measure ANOVA yielded a significant main effect of speech task on vocal intensity [$F(5, 60) = 23.92$, $p < 0.001$]. Task $\times$ group interactions and the main effect of group on vocal intensity were non-significant. Pairwise comparisons revealed significantly greater vocal intensity during loud speech compared to all other speech conditions ($p \leq 0.001$). Further, slow speech was produced with significantly less vocal intensity than all other speech conditions ($p < 0.02$) and fast speech was produced with significantly more vocal intensity than typical speech ($p < 0.03$).

IV. DISCUSSION

This study aimed to determine the associative strength between tongue movement pattern consistency and formant movement patterns. Younger and older adults were asked to repeat one sentence under various speech conditions known to elicit changes in movement pattern consistency to observe corresponding changes in formant movement patterns. Outcomes revealed that 85% of the variability in formant movement pattern consistency across talkers at typical speech was accounted for by tongue movement pattern consistency. However, two talkers who had extremely high STI values in the kinematic and acoustic domain heavily influenced the strength of the relation. Further, 41% of the

FIG. 6. Means ($\pm$ SE) of utterance durations (top) and vocal intensities (bottom) of younger and older adults across all speech tasks.

variability in formant movement pattern consistency across various speech tasks was accounted for by tongue movement pattern consistency.

A. Associations between tongue movement and formant movement pattern consistency across younger and older talkers

Younger and older adults were enrolled in this study to elicit a wide performance range in movement pattern consistency. Based on previous studies, it was expected that older adults would produce less consistent tongue movement patterns (higher tongue STI values) than younger adults (e.g., Wohlert and Smith, 1998; Kleinow *et al.*, 2001). However, in the present study only two older adults showed noticeably higher tongue STI values compared to all other talkers during typical speech. Sadagopan and Smith (2013) also reported no significant differences in the lip kinematic STI between younger and older adult talkers for pseudoword productions. These researchers suggested that the age range of the participants was too low (66–74 years of age). In the present study, participants in the older age group were between 75 and 92 yr old, which was similar to the age range of the study of Wohlert and Smith (1998). In the future, it may be useful to also consider factor such as resting heart rate and forced vital capacity (Ramig and Ringel, 1983) in addition to chronological age to elicit a larger range in speech motor performance.

Nevertheless, the formant acoustic STI values of those two older adults coincided with the elevated tongue STI values. This finding provides preliminary support for the notion that changes in tongue movement pattern consistency yield

predictable effects on formant movement patterns. However, the strength of association that was observed for articulatory-acoustic relations across talkers may be inflated because only two older adults exhibited extremely elevated STI levels. Therefore, the findings of relatively high predictability of tongue movement pattern consistency based on formant movements across talkers should be treated with caution. Future studies are needed to further examine articulatory-acoustic relations across talkers. These studies should include talkers with impaired speech movements who are known to have highly inconsistent movements patterns (e.g., talkers with dysarthria post traumatic brain injury; see McHenry, 2003).

With regard to the tongue motor performance of older adults, findings of this study suggest that the aging effect on speech motor control may not be as pronounced as previously thought (e.g., Kleinow *et al.*, 2001). It is also possible that an aging-related decline in tongue motor control is more evident in female talkers than in male talkers because the two talkers who exhibited the notably higher tongue kinematic and formant acoustic STI values were two of the three females enrolled in the study (neither of them wore hearing aids). Trends of such sex-specific aging effects on speech production have been reported previously by Weismer and Fromm (1983) for the intrasubject variability of voiceless interval durations for voiceless fricative or stop production although significance was also not reached due to a small sample size.

B. Associations between tongue movement and formant movement pattern consistency across speech behavioral modifications

In general, speech task effects on formant movement pattern consistency coincided with speech task effects on tongue movement pattern consistency. The moderate association between tongue kinematic and formant acoustic STI across tasks suggests that task-related changes in tongue movement pattern consistency are reflected in the formant movement patterns. This finding is in contrast to previous findings based diphthong transitions where no significant associations between tongue kinematic and formant acoustic movement patterns were observed (Mefferd and Green, 2010). One likely reason for difference in articulatory-acoustic mappings may be the larger range in task-related performance change in tongue kinematic movement consistency that was elicited in the current study compared to the previous one, which used diphthong transitions to calculate kinematic and acoustic STIs. Thus, to study tongue motor control indirectly a complex movement signal, such as those elicited by phrase- or sentence-level utterances, is advisable to elicit sufficient range in task effects.

Slow, fast, and loud speech effects on tongue movement pattern constituency were in congruence with those reported in previous lip and jaw kinematic studies (e.g., Smith *et al.*, 1995; Kleinow *et al.*, 2001; McHenry, 2003; Mefferd *et al.*, 2014). Although these task effects on movement pattern variability are considered relative robust for jaw and lip movements, to this day they have rarely been examined in tongue

kinematic studies. Taking together previous findings and findings from this study, task-dependent changes in movement pattern consistency appear to be similar across articulators. This similarity across articulators suggests that task-related changes in movement pattern consistency may be attributed to central motor control processes rather than peripheral (e.g., muscle-specific, biomechanical) mechanisms.

The effect of clear speech and bite block speech on speech movement pattern consistency has not been studied much. For clear speech, one study has reported decreased lip movement pattern consistency in talkers with Parkinson's disease (Dromey, 2000). The current study did not concur with these findings, which may suggest that clear speech effects differ between impaired and typical speakers. Further, bite block speech, which is commonly thought to perturb the typical execution of speech movements because tongue and lip movements have to be adjusted to accommodate for the lowered and fixated jaw position, appeared to elicit a decrease in tongue movement patterns. However, when looking more closely at each individual's data, tongue movement pattern consistency only decreased in two participants (Y1 and Y3) in response to the bite block speech. Thus, for the majority of participants, speaking with a bite block and moving the tongue independently of the jaw did not negatively affect tongue movement pattern consistency. This finding is in agreement with previous observations of bite block speech and lip movement pattern consistency (Namasivayam *et al.*, 2008).

C. Individual articulatory-acoustic relations for movement pattern consistency

Whereas for many talkers in the younger and older group, articulatory-acoustic relations were strong, for some talkers articulatory-to-acoustic relations were virtually nonexistent. When articulatory-acoustic relations were weak, the selected speech tasks commonly elicited only a small range of tongue kinematic STI values. This finding is concurred with our previous study, where a small range in tongue kinematic STI values based on the diphthong also resulted in weak articulatory-acoustic relations for movement pattern consistency (Mefferd and Green, 2010).

As shown in Table I, clear speech had a negative impact on the overall strength of articulatory-acoustic relations for one talker (Y6), whereas the bite block condition decreased the strength of relation between articulatory and acoustic movement pattern consistency in three other talkers (Y1, Y3, O8). The negative impact of these conditions may be explained by a change in articulatory strategy to achieve the desired speech acoustic signal. Specifically, talkers may trade-off articulatory movement of the posterior tongue with movement of other areas in the vocal tract (e.g., pharyngeal and laryngeal adjustments). Such articulatory tradeoffs in response to perturbed speech conditions have been observed previously for tongue-body and lip rounding (Perkell *et al.*, 1993), upper and lower lip (e.g., Abbs *et al.*, 1984), as well as lip rounding and laryngeal height adjustments (Riordan, 1977). In the present study, Y3 demonstrated a considerable reduction in posterior tongue displacements during the bite block condition,

suggesting articulatory trade-offs with other articulatory structures. By contrast, posterior tongue movements of Y1, Y6, and O8 did not show considerable changes from typical speech in their displacements. These talkers may have moved other parts of the vocal tract in addition to the posterior tongue to produce the target utterance and varied these adaptive strategies across repetitions, which would explain the relatively low tongue movement pattern consistency in combination with high formant movement pattern consistency. In future studies, the examination of more than one fleshpoin within the vocal tract may provide better insights in such adaptive strategies, which should result in a more comprehensive understanding of articulatory-acoustic relations.

D. Future work

Based on the talkers included in this study, the potential of the acoustic STI measure as an indirect acoustic measure of tongue motor control is obvious. Its sensitivity, however, has to be further explored. Future studies could also expand on the findings outside of the typical performance range by enrolling talkers with highly variable speech movement patterns, such as talkers with dysarthria post traumatic brain injury (McHenry, 2003). Based on preliminary findings, it is expected that abnormally high tongue kinematic STI values should be reflected in the speech acoustic domain as abnormally high formant acoustic STI values.

Finally, it should be mentioned that the acoustic analysis, although automated by software-based LPC-tracking is very time-consuming and requires trained skills to recognize and hand-correct formant tracking errors. This could potentially be a barrier for the use of the acoustic STI as an objective assessment tool for tongue motor control in clinical settings. However, the acoustic STI may be a viable proxy measure in research settings where tongue kinematic data are too expensive or impossible to acquire (e.g., small children, talkers for whom tongue sensor placements are too difficult).

ACKNOWLEDGMENTS

This work was supported in part by the American Speech Language and Hearing Foundation's New Investigators Grant and NIH Grant No. RDC015075A, as well as start-up money provided by Wichita State University and Vanderbilt University Medical Center. Preliminary results were presented in March 2014 at the Conference on Motor Speech in Sarasota, FL. We would like to thank all participants for their willingness to complete this study. The author would like to thank Emma Cerre for many, many hours of formant tracking and Nicholas Coates for his assistance with the MATLAB scripts for the formant interpolation.

- Abbs, J., Gracco, V., and Cole, K. (1984). "Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming," *J. Mot. Behav.* **16**(2), 195–231.
- Adams, S. G., Weismer, G., and Kent, R. D. (1993). "Speaking rate and speech movement velocity profiles," *J. Speech Hear. Res.* **36**(1), 41–56.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 77–81.

- Cummins, F., Lowit, A., and van Brenk, F. (2014). "The quantitative assessment of inter-utterance stability: Application to dysarthria," *J. Speech, Lang. Hear. Res.* **57**, 81–89.
- de Boor, C., Höllig, K., and Sabin, M. (1987). "High accuracy geometric hermite interpolation," *J. Aided. Geom. Des.* **4**, 269–278.
- Dromey, C. (2000). "Articulatory kinematics in patients with Parkinson's disease using different speech treatment approaches," *J. Med. Speech-Lang. Pathol.* **8**, 155–161.
- Dromey, C., and Ramig, L. (1998). "Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation," *J. Speech, Lang. Hear. Res.* **41**, 1003–1018.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague, Netherlands).
- Folstein, M. F., and Folstein, S. E. (2012). *Mini-Mental State Examination*, 2nd ed. (PAR, Inc., Lutz, FL).
- Gay, T., Lindblom, B., and Lubker, J. (1981). "Production of bite-block vowels: Acoustic equivalence by selective compensation," *J. Acoust. Soc. Am.* **69**, 802–810.
- Green, J. R., Moore, C. A., and Reilly, K. J. (2002). "The sequential development of lip and jaw control for speech," *J. Speech, Lang. Hear. Res.* **45**, 66–79.
- Green, J. R., Wang, J., and Wilson, D. L. (2013). "SMASH: A tool for articulatory data processing and analysis," *Interspeech*, pp. 1331–1335.
- Grigos, M. I. (2009). "Changes in articulator movement variability during phonemic development: A longitudinal study," *J. Speech, Lang. Hear. Res.* **52**(1), 164–177.
- Grigos, M. I., Moss, A., and Lu, Y. (2015). "Oral articulatory control in childhood apraxia of speech," *J. Speech, Lang. Hear. Res.* **58**(4), 1103–1118.
- Hausdorff, J. M. (2007). "Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking," *Human Movement Sci.* **26**(4), 555–589.
- Hausdorff, J. M., Lertratanakul, A., Cudkowicz, M. E., Peterson, A. L., Kaliton, J. D., and Goldberger, A. L. (2000). "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol.* **88**(6), 2045–2053.
- Howell, P., Anderson, A., Bartrip, J., and Bailey, E. (2009). "Comparison of acoustic and kinematic approaches to measuring utterance-level speech variability," *J. Speech, Lang. Hear. Res.* **52**, 1088–1096.
- Kleinow, J., Smith, A., and Ramig, L. O. (2001). "Speech motor stability in IPD: Effect of rate and loudness manipulations," *J. Speech, Lang. Hear. Res.* **44**, 1041–1051.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H and H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–440.
- Lindblom, B., and Sundberg, J. (1971). "Acoustic consequences of lip, tongue, jaw, and larynx movements," *J. Acoust. Soc. Am.* **50**, 1166–1179.
- McHenry, M. A. (2003). "The effect of pacing strategies on the variability of speech movement sequences in dysarthria," *J. Speech, Lang. Hear. Res.* **46**, 702–710.
- Mefferd, A., and Green, J. R. (2010). "The articulatory-to-acoustic relationship in response to speaking rate and loudness manipulations," *J. Speech, Lang. Hear. Res.* **53**(5), 1206–1219.
- Mefferd, A. S., Pattee, G. L., and Green, J. R. (2014). "Speaking rate effects on articulatory pattern consistency in talkers with mild ALS," *Clin. Linguist. Phon.* **28**(11), 799–811.
- Milenkovic, P. (2005). "Time-frequency analysis software program for 32-bit Windows [computer software]," <http://userpages.chorus.net/cspeech/> (Last viewed 11/3/2016).
- Murdoch, B., Chen, H. Y., and Goozee, J. V. (2012). "Developmental changes in the variability of tongue and lip movements during speech from childhood to adulthood: An EMA study," *Clin. Linguist. Phon.* **26**(3), 216–231.
- Namasivayam, A., van Lieshout, P., and de Nil, L. (2008). "Bite-block perturbation in people who stutter: Immediate compensatory and delayed adaptive processes," *J. Commun. Disord.* **41**, 372–394.
- Netsell, R. (1985). "Construction and use of a bite-block for the evaluation and treatment of speech disorders," *J. Speech Hear. Disord.* **50**(1), 103–106.
- Newell, K., and Corcos, D. (1993). "Issues in variability and motor control," in *Variability and Motor Control*, edited by K. Newell and D. Corcos (Human Kinetics, Champaign, IL), pp. 1–12.
- Nip, I. S. B., Green, J. R., and Marx, D. B. (2009). "Early speech motor development: Cognitive and linguistic considerations," *J. Commun. Disord.* **42**, 286–298.
- Perkell, J., Matthies, M., Svirsky, M., and Jordan, M. (1993). "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study," *J. Acoust. Soc. Am.* **93**(5), 2948–2961.
- Ramig, L., and Ringel, R. (1983). "Effects of physiological aging on selected acoustic characteristics of voice," *J. Speech Hear. Res.* **26**, 22–30.
- Riordan, C. J. (1977). "Control of vocal-tract length in speech," *J. Acoust. Soc. Am.* **62**(4), 998–1002.
- Sadagopan, N., and Smith, A. (2008). "Developmental changes in the effects of utterance length and complexity of speech movement variability," *J. Speech, Lang. Hear. Res.* **51**, 1138–1151.
- Sadagopan, N., and Smith, A. (2013). "Age differences in speech motor performance on a novel speech task," *J. Speech, Lang. Hear. Res.* **56**, 1552–1566.
- Sasisekaran, J., Smith, A., Sadagopan, N., and Weber-Fox, C. (2010). "Nonword repetition in children and adults: Effects on movement coordination," *Dev. Sci.* **13**(3), 521–532.
- Schötz, S., Frid, J., and Löfqvist, A. (2013). "Development of speech motor control: Lip movement variability," *J. Acoust. Soc. Am.* **133**(6), 4210–4217.
- Schmidt, R. (2003). "Motor schema theory after 27 years: Reflections and implications for a new theory," *Res. Q. Exerc. Sport* **74**(4), 366–375.
- Schmidt, R., and Lee, T. (2005). *Motor Control and Learning: A Behavioral Emphasis*, 4th ed. (Human Kinetics, Champaign, IL).
- Schmidt, R., and Wrisberg, C. (2004). *Motor Learning and Performance*, 3rd ed. (Human Kinetics, Champaign, IL).
- Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., and McGillem, C. (1995). "Spatiotemporal stability and patterning of speech movement sequences," *Exp. Brain Res.* **104**, 493–501.
- Stergiou, N., and Decker, L. M. (2011). "Human movement variability, nonlinear dynamics, and pathology: Is there a connection?," *Human Movement Sci.* **30**(5), 869–888.
- Stevens, K. N. (1972). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, edited by P. B. Denes and E. E. David, Jr. (McGraw-Hill, New York), pp. 51–66.
- Stevens, K. N. (1989). "On the quantal nature of speech," *J. Phon.* **17**, 3–45.
- Sjölander, K., and Beskow, J. (2011). "Wavesurfer 1.8 [software program]" (Kungliga Tekniska Högskolan, Sweden).
- Tjaden, K., and Wilding, G. (2004). "Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings," *J. Speech, Lang. Hear. Res.* **47**, 766–783.
- Walsh, B., Smith, A., and Weber-Fox, C. (2006). "Short-term plasticity in children's speech motor systems," *Dev. Psychobiol.* **48**(8), 660–674.
- Weismer, G., and Fromm, D. (1983). "Acoustic analysis of geriatric utterances: Segmental and nonsegmental characteristics that relate to laryngeal function," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College Hill, San Diego), pp. 317–332.
- Wohler, A. B., and Smith, A. (1998). "Spatiotemporal stability of lip movements in older adult speakers," *J. Speech, Lang. Hear. Res.* **41**(1), 41–50.
- Yunusova, Y., Green, J. R., and Mefferd, A. S. (2009). "Accuracy assessment for the AG500, Electromagnetic Articulograph," *J. Speech, Lang. Hear. Res.* **52**(2), 547–555.