1720

The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance

Khalil Iskarous[a)]
Department of Linguistics, University of Southern California, Grace Ford Salvatori 301, Los Angeles, California 90089-1693

Christine Mooshammer[b)]
Department of German Language and Linguistics, Humboldt University, Unter den Linden 6, D-10099 Berlin, Germany

Phil Hoole
Institut für Phonetik und Sprachverarbeitung, LMU München, Schellingstrasse 3, D-80799 München, Germany

Daniel Recasens
Departament de Filologia Catalana, Universitat Autònoma de Barcelona, 08913 Bellaterra, Barcelona, Spain

Christine H. Shadle
Haskins Laboratories, 300 George Street, Suite 900, New Haven, Connecticut 06511

Elliot Saltzman
Boston University, 675 Commonwealth Avenue, Boston, Massachusetts 02215

D. H. Whalen
The Graduate Center, CUNY, 365 Fifth Avenue, New York, New York 10016

(Received 17 February 2013; revised 29 April 2013; accepted 17 June 2013)

Coarticulation and invariance are two topics at the center of theorizing about speech production and speech perception. In this paper, a quantitative scale is proposed that places coarticulation and invariance at the two ends of the scale. This scale is based on physical information flow in the articulatory signal, and uses Information Theory, especially the concept of mutual information, to quantify these central concepts of speech research. Mutual Information measures the amount of physical information shared across phonological units. In the proposed quantitative scale, coarticulation corresponds to greater and invariance to lesser information sharing. The measurement scale is tested by data from three languages: German, Catalan, and English. The relation between the proposed scale and several existing theories of coarticulation is discussed, and implications for existing theories of speech production and perception are presented.
 [http://dx.doi.org/10.1121/1.4812855]

PACS number(s): 43.70.Bk, 43.70.Jt, 43.70.Mn [CYE] Pages: 1271–1282

I. INTRODUCTION

Coarticulation, the influence of one phonetic segment on the realization of another is a major cause of articulatory and acoustic variability in speech (Lindblom, 1963). But despite the prevalence of coarticulatory variability in speech, many scientists have proposed invariants at the acoustic, articulatory, auditory, and cognitive levels (Stevens and Blumstein, 1978; Fowler, 1998; Syrdal, 1985; Lahiri and Marslen-Wilson, 1991). Most researchers agree that there is extensive coarticulatory overlap in speech, but also that some aspects of production and acoustic output for each segment may be more invariant than other aspects, namely those aspects most crucial for the achievement of the segment. One of the sources of theoretical disagreement is the lack of a unified quantitative approach to allow researchers to determine when a particular aspect of the speech process is invariant, or whether it exhibits coarticulatory variability. In this paper we propose the Coarticulation/Invariance Scale, a quantitative scale having invariance on one end and coarticulation on the other. The measure relies on characterizing physical information flow in the articulatory signal, using mutual information (MI) as a quantitative measure. MI is used to measure the amount of information sharing between contiguous phonological units, or segments, and will be shown to be large under coarticulation and small for aspects of the articulation of a segment that are relatively invariant. The information flow measured is *physical* in that it relies on the measurement of the physical positions of articulators in speech production. This method is highly related to the method of locus equations, used extensively for measuring

[a)]Author to whom correspondence should be addressed. Electronic mail: kiskarou@usc.edu
[b)]Also at: Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511.

coarticulation acoustically (Sussman *et al.*, 1991; Lindblom and Sussman, 2012; Iskarous *et al.*, 2010). The relation between the new scale and locus equations, as well as other theories of coarticulation will be addressed.

Most studies on coarticulation and invariance measure speech in terms of the spatial and temporal physical units that current laboratory instruments quantify: millimeters, Hertz, Barks, milliseconds, mm/s, etc. The hypothetical approach we test in this paper is more abstract and quantifies information flow within speech across time, based on spatiotemporal measurements. Coarticulation is measured in this approach as presence of information about a segment B in a time period when another segment A is being produced. Invariance, on the other hand, is measured as the absence of information about other segments, when a segment A is being produced.

We hypothesize that coarticulation and invariance are on two ends of a scale. We believe that the reason for this is the well-known phenomenon of coarticulation resistance. It has been known for a long time that different segments influence and are influenced by surrounding segments to different extents (Bladon and Al-Bamerni, 1976). Segment-to-segment difference in resistance to coarticulation is thought to be due to differences in how the vocal tract articulators are constrained in each segment (Recasens, 2012; Recasens and Espinosa, 2009; Mooshammer *et al.*, 2006). Moreover, it has been demonstrated that place of articulation and manner of articulation have effects on the coarticulation resistance. For instance, alveo-palatals have higher coarticulation resistance than alveolars (Recasens and Espinosa, 2009), and coronal sibilants have higher coarticulation resistance for the jaw than non-sibilant coronals (Mooshammer *et al.*, 2006). This is exactly the scalar difference we seek to measure using the coarticulation/invariance scale. The scale we propose builds on three previous approaches: degree of articulatory constraint (DAC) model (Recasens and Espinosa, 2009), locus equations model (Sussman *et al.*, 1991; Lindblom and Sussman, 2012), and the Jackson-Singampalli Statistical Identification model (Jackson and Singampalli, 2009).

DAC seeks to quantify the difference between segments in how they resist coarticulation and how they impose their articulatory demands on neighboring segments. The quantification works by first measuring, in physical units, for each segment, the degree to which the segment influences and is influenced by its surrounding segments. Numbers are then assigned to each segment by the investigator that reflect the relative resistance of the segments. For instance, sibilants have a higher DAC index than labials, since the latter involve far fewer constraints on the tongue than the former. The DAC index is empirically derived, in that it is based on physical measurements of the extent of coarticulation, but it is an abstract *qualitative* measure of the resistance itself. DAC has been applied mostly to coarticulation resistance due to the tongue and jaw's articulatory constraints. However, the lips for /p/ can be resistant to certain lip activity of surrounding vowels, even if its tongue activity is not resistant to those same vowels. The measure used in this work will generalize DAC so that it is based on all measured articulators, not just the tongue and jaw. Moreover, the measure derived in this work is not assigned qualitatively; it is quantitatively derived from speech data.

One method of quantifying of coarticulation resistance, defined in the acoustic domain, is provided by locus equations (LE) (Lindblom, 1963; Sussman *et al.*, 1991; Brancazio and Fowler, 1998), which are regression lines estimated by predicting F2 of a consonant from F2 of a vowel, whereby the consonant is fixed and the vowel varies over the possible vowels of the language. The slopes and intercepts of these lines vary systematically as place of articulation of the consonant changes. Krull (1987, 1989) and Fowler (1994) showed that LE are a measure of coarticulation degree, with the slope increasing with coarticulation degree, which is itself inversely correlated to coarticulation resistance. Specifically, labials have the highest LE slope, highest coarticulation degree, and lowest coarticulation resistance, whereas alveolars have the lowest LE slope, lowest coarticulation degree, and highest resistance. Iskarous *et al.* (2010) showed that LE do measure coarticulation resistance, but only those aspects of resistance that are contributed by the tongue back's horizontal motion. Therefore DAC indices are similar to LE in privileging the tongue over the rest of the vocal tract articulators in trying to determine a segment's resistance level. However, LE, unlike DAC indices, are quantitatively derived from data. Iskarous *et al.* (2010) showed that LE's quantification of coarticulation resistance is based on the statistical concept of predictability, via linear regression. The more predictable a consonant's articulation and acoustics are from a contiguous vowel (as the vowel varies), the lower the coarticulation resistance of the consonant, and the lower the LE slope. In this paper, we propose to extend the quantification of coarticulation resistance initiated in the LE literature in two ways. First we use mutual information, a general measure of independence that is insensitive to the probability distribution of the data. Second we apply this general statistical independence measure in the articulatory domain to determine for each consonant, the degree of predictability of each articulator's position from the position of a contiguous vowel's articulator position, as the vowel varies. Therefore the methods developed here should allow the generalization of previous results to several articulators (e.g., jaw, lips, and other points of the tongue).

A statistical method for measuring coarticulation resistance was proposed by Recasens (1985) and Recasens and Espinosa (2009). In this method, the range or standard deviation of a measured variable (such as formant value or articulator position) for a segment is used as an indication of how constrained that variable is in the production of the segment. Small variability for a segment means that the production of that variable is highly constrained for the segment, whereas a large standard deviation means that the production of that variable is unconstrained for the segment. Therefore the smaller the variability, the greater the resistance. The problem for such an analysis, however, is what to use as a standard for high or low for the standard deviation or range. Two methods have been proposed to solve this problem, the Jackson-Singampalli Statistical Identification model and the MI model, to be described below. Jackson and Singampalli (2009) propose an algorithm for determining the critical, dependent, and redundant articulators for a segment based on previously collected articulatory data. In this method, the distribution of each articulator position at the middle of segments is estimated from an entire database of movement data. Then the distribution of positions for that same articulator is extracted for each segment, separately; these are the phone distributions. The grand and phone distributions are fitted with Gaussian distributions, and then the Kullback-Leibler divergence between each phone distribution and the grand distribution is measured. The critical articulators for a segment are those that show a large divergence between phone and grand distributions. This method generalizes Recasens' range/standard deviation measure by taking the entire distribution of the variable into account, not just the variability of the distribution. Moreover the distribution is compared to the grand distribution, establishing a criterion and a scale (a divergence scale) for comparing phone distributions. The method is more complex however, since many articulators could show a large divergence from the grand distribution, due to the correlation of movement of a non-critical with the critical articulator. The method therefore involves an additional iterative algorithm for computing inter-articulator correlations and adjusting iteratively for these correlations. Intuitively, the method works because the grand distribution for an articulator, in one or two dimensions, shows how the position of the articulator can vary across all segments. Phones that have a similar distribution to the grand distribution are ones that make no greater demands of the articulator than the demands made by an average segment, whereas phones that have a distinct distribution are likely to make special demands of that articulator, indicating that it is a critical articulator for that phone.

A different way to assess the criticality of an articulator for a particular segment A is to try to predict the position of the articulator in segment A from the following segment x's articulator position, as that latter segment x varies (locus equation paradigm). If segment A's position for the articulator is highly predictable from following segment x, then that articulator is not critical for the segment A, since the predictability indicates that the articulator's position for A is not determined by A itself, but by the following context x. However, if segment A's position for the articulator is independent of the following segment x, then it means that the articulator is critical for segment A, since segment A determines its own value for the position of the articulator. Therefore criticality of an articulator for a segment A could be measured by the predictability vs lack of predictability of the position of the articulator from context x. The idea we seek to explore in this paper is that Mutual Information, a measure of predictability described in the next section, can be used to define a quantitative scale with invariance at the low end and coarticulation at the high end, and that that measure is sensitive to coarticulation resistance. We test this hypothesis by measuring coarticulation in four datasets, two from German, one from Catalan, and one from American English, using a measure of mutual information for each articulator of several consonants and vowels. These datasets have been previously analyzed in the literature on coarticulation resistance, so we are able to compare mutual information measures with published results based on these same datasets. The analyses from each of the datasets allow us to cover different aspects of how MI quantifies coarticulation. The first German dataset consists of the three voiceless stops /p,t,k/, allowing us to determine if effects of place of articulation on coarticulation are quantifiable. The second German dataset consists of coronals with different manners of articulation, allowing us to determine if effects of manner on coarticulation are resolvable. The Catalan data, which includes alveopalatal and alveolar nasals in the same contexts allows us to determine the effects of place differences within nasal manner on coarticulation. The English dataset is based on a single consonant /s/ in different vowel contexts, but MI is measured from the beginning to the end of /s/, allowing for the temporal evolution of coarticulation to be investigated.

II. MUTUAL INFORMATION

Mutual information is a general measure of independence between two variables, which was developed within Information Theory (Shannon, 1948) to determine the degree to which two sources of information (variables) share information. The elementary measure of information within Information Theory is *entropy*, a measure of how uncertain the possible outcomes of measurement of a variable are. If a variable can take on any possible value, it has a uniform probability distribution, and maximal entropy: if its outcome is limited to one possible value, it has zero entropy. Shannon quantified entropy as

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i), \tag{1}$$

where H is the entropy, X is the random variable, and $p(x_i)$ is the probability of the occurrence of measurement x_i for variable X.

When there are two sources of information, i.e., two variables, it is possible to measure the shared information between them by quantifying the degree to which knowledge of the outcome of one variable limits the possibilities of the other variable. If the two variables are independent, i.e., have zero MI, then knowledge of the outcome of one variable tells us nothing about the possible outcomes of the other, whereas a non-zero MI means that knowledge of the outcome of one variable does indeed limit the possible outcomes of the other. In probability theory, the joint probability of two independent events is simply the product of those two probabilities (e.g., the probability of obtaining two heads in two fair-coin tosses is $\frac{1}{4}$, since the probability of a head is $\frac{1}{2}$ for each fair coin). It is therefore possible to measure the departure from independence between two variables by comparing their actual joint distribution with the joint distribution based on the assumption of independence. MI is a quantification of independence through a comparison of these two probability distributions, the measured joint distribution and the joint distribution assuming independence

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}, \tag{2}$$

where MI(X, Y) is the mutual information of the variables X and Y; $p(x, y)$ is the measured joint distribution of the variables, and $p(x)p(y)$ is the joint probability distribution assuming that X and Y are independent.

In this paper, MI of consonants is calculated by choosing a consonant, e.g. /p/, and a component of the position of one of the articulators, then measuring the consonant's

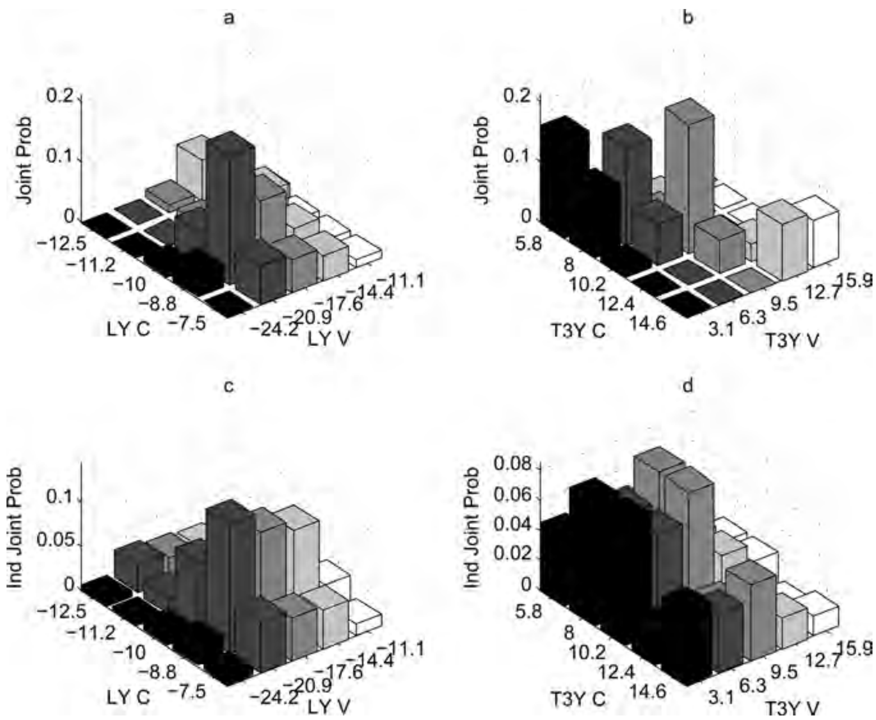

FIG. 1. Example of joint probability distributions based on German data of /p/ for one speaker, as discussed in Sec. IV A, as inputs to MI calculation for the vertical component of the lips [left panels (a),(c)] and tongue predorsum [right panels (b),(d)] The upper panels (a),(b) are histograms computed directly from the paired V and C data, while the lower panels (c),(d) are the joint probability distributions for V and C, based on the assumption of independence. The horizontal and vertical scales are in millimeters.

articulator position as a function of the position of that same articulator for different vowels in the language. For instance, the *y*—component of the lower lip pellet, L_y is measured in the middle of /p/ before different vowels of the language. Then we ask how much the articulator's position at the temporal midpoint of the consonant can be predicted from the same articulator's measurement at the midpoint of the vowel, as the vowel varies. Here, C and V are the two sources of information, or variables, and we can calculate the MI, as the shared information. This is done by calculating the joint distribution of C and V, and comparing this joint distribution to the joint distribution that is based on the assumption that V and C are independent, using Eq. (2). Figure 1 provides two examples of MI calculation for /p/ in German. The position scales are in millimeters, and their range corresponds to the range of variability for one speaker. The data were divided into five bins. Panel (a) shows the joint probability distribution for a fixed C for the vertical component of the lips; panel (b) shows the joint distribution for T3y for the same subject. Below each, we see the joint distributions assuming V and C independence. It can be seen from the figure that the MI joint distribution for L_y (a) looks quite similar to the independent joint distribution in (c); the MI calculated is 0.165, the lowest value for the subject. In contrast, (b) looks quite different from the independent joint distribution in (d), and the MI calculated is much larger, 0.78, indeed the highest value for the subject. Low MI, therefore indicates little to no shared information, i.e., independence, whereas high MI indicates dependence.

One of the main decisions involved in making an MI measurement is the number of bins to include in the estimation of the histogram function, the most naive estimator of probability distribution. The distribution is represented more faithfully by a larger number of bins, but more bins require more data. In this paper, we used three bins, meaning that only the coarsest aspects of the probability distributions are preserved. One additional decision commonly occurring in Mutual Information estimation from sparse data is that a bin in a histogram with no entries, and thus producing an estimate of zero probability, can cause MI to become infinite. There are several long-standing methods for fixing this problem. We used the Jeffrey-Perks law, which adds 0.5 to all cells prior to MI estimation. The effect on MI estimation of increasing bin width is investigated in the Sec. IV.

III. METHODS

As mentioned above we present results from four corpora. The first corpus was designed to investigate the German vowel system. Tongue, jaw and lip movements from seven speakers of Standard German (six male, one female) were recorded by means of 2D electromagnetic midsagittal articulography (EMMA AG100, Carstens Medizinelektronik AG). Four sensors were placed on the midline of the tongue from 1 to 6 cm behind the tongue tip (T1) to the back of the tongue (T4) and two sensors in-between (T2, T3). One sensor was placed on the gum below the lower incisors to monitor jaw movements and one sensor on the skin just below the lower lip. Two sensors on the upper incisors and the nasion were used for head movement correction. The sampling rate of the movement data was 250 Hz [for further details on post-processing of this data set, see Hoole (1999)]. The material consisted of symmetrical CVC sequences with the voiceless stops /p, t, k/ and the 15 stressed vowels /iː yː eː øː ɛː aː oː uː ɪ ʏ ɛ œ a ɔ u/. These sequences were produced in the sentence Ich habe geCVCe gesagt. (I said geCVCe.) Each sentence was repeated five times in random order and the whole corpus, consisting of 225 sentences (3 stops × 15 vowels × 5 repetitions), was produced at both a normal and a fast speaking rate. For calculating MI the sensor positions at the midpoints of the initial consonant, the vowel and the final consonant were used. The midpoints were determined using acoustic labels for each speech sound. For the consonants the midpoint was measured halfway through the acoustically defined closure. The onset of the vowel was not labeled in this dataset; therefore, for the vowel the interval from the burst of the initial stop to the acoustic offset of the vowel was measured, and the point one-third from the end of the vowel was treated as the vowel midpoint. Sec. V C 2 presents locus equations analysis based on this data. Slope and explained variability R^2 are calculated via least squares.

The second German corpus was designed to investigate the role of the jaw for the production of coronal consonants in German. One female and four male speakers of standard German were recorded by means of EMMA. Sensor placement was the same as above except that one sensor was placed on the chin instead of the lower lip and an additional sensor was placed on the inside of the lower incisors. The six coronal consonants /s, ʃ, t, d, n, l/ were recorded in symmetrical VCV sequences between the vowels /iː eː aː/. All VCV sequences were embedded in the carrier phrase "Hab das Verb VCV mit dem Verb VCV verwechselt" (I mixed up the Verb VCV with the Verb VCV). Each sequence was repeated six times in the first position and six times in the second position in randomized order. The whole corpus was recorded at both, a comfortable volume and in loud (but not shouting) speech. Measurements of the tongue and jaw sensor positions were taken from the acoustically defined midpoint of the first vowel, the consonant and second vowel (for further details on post-processing, see Mooshammer *et al.* (2006).

The Catalan data set consists of articulatory movement data from three male native speakers of Eastern Catalan recorded by means of a 2D EMMA (AG100, Carstens Medizinelektronik AG). Three sensors were attached midsagittally to the tongue (tongue tip = T1, tongue dorsum = T3, in-between = T2), and one sensor each to the lower lip (LL), the upper lip (UL), and the lower incisors (J). The material analyzed in this paper consisted of symmetrical /pVCVp/ sequences with the vowels /i, a, u/ and the consonants /n, ɲ/. MI was calculated based on V1. The test words were equally stressed on the first and the second syllable and embedded in a meaningful carrier phrase that was repeated 10 times. Tongue, jaw, and lip positions were extracted at the midpoint of the acoustically defined consonant and vowel.

The American English data for examining the time course of coarticulation during /s/ are from the x-ray microbeam database (XRMB) (Westbury, 1994, pp. 88–107). The particular data examined are from Task 13, a list of /sVd/ words: side, sewed, seed, sod, sued, sawed, sid, sad, surd, said. Data from 24 subjects are included (15 females, nine males). In this paper, we analyze the data for the vertical and horizontal components of the lips (LLx and LLy) and the vertical components for the jaw and tongue tip (Jy and T1y). The /s/ and vowels were acoustically segmented, as described in Iskarous *et al.* (2011). The articulator time series for /s/ were then uniformly extracted at 10 points, and the articulator positions at the midpoint of the vowels were extracted. MI was calculated for each of the 10 points during the /s/, with the vowel's contribution always being from its midpoint. Further details on the dataset can be found in Iskarous *et al.* (2011).

A. Statistics

The MI data will be qualitatively described by referring to figures showing MI as a function of segment and articulator. Statistical tests were all mixed-effects general linear models, with Subject as the random effect. Significance is determined through a Markov Chain Monte Carlo simulation. All statistical tests were performed through the R program **lmer** (Baayen, 2007, pp. 241–300).

IV. RESULTS

In the following subsections, we present MI measurements on our four datasets. The first German dataset allows us to investigate the relationship between place of articulation and MI; the second, the relationship between manner of articulation and MI. The Catalan dataset allows for the examination of additional place distinctions, and the English dataset allows us to examine how MI reveals information about the time course of coarticulation. Since the scale proposed here is related to research on locus equations, we present locus equations measurements on the first German dataset. This is followed by an examination of how bin width affects MI measurement.

A. German stop data: Place of articulation effects on MI

Our first analysis tested whether MI can be used to distinguish between different places of articulation by examining MI as a function of segment and articulator for labial, alveolar, and velar stops. Figure 2 shows MI for the German stops /p,t,k/ (crosses, diamonds, and circles, respectively) for seven subjects. Panel (a) presents MI measured for the vertical components of the articulators indicated, and panel (b) for the horizontal components. Variability is over subjects. These data show that for the lip pellet (L), /p/ has lower MI relative to its adjacent vowel than /t/ or /k/, but the effect is greater in the vertical than the horizontal components. For the tongue tip pellet (T1) /t/ has lower MI than /p/ and /k/ in

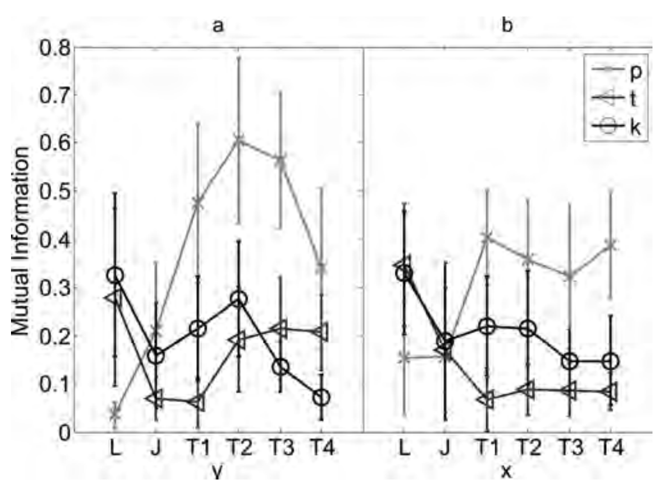

FIG. 2. Mutual information as a function of articulator for three German stops for seven subjects. The error bars are centered at the mean and span a standard deviation on each side of the mean. The data are divided into vertical (a) and horizontal (b) components. L is lower lip; J is jaw; T1 is the most anterior tongue sensor and T4, the most posterior.

both components. For the most posterior pellet on the tongue dorsum, T4, /k/ has a lower MI than /p/ and /t/, but only for the vertical components. For the jaw pellet (J), there is a great deal of overlap between stops: for the tongue pellets, /p/ in general has higher MI than the lingual consonants.

In the quantitative tests, the dependent variable was MI for either the vertical or horizontal component of an articulator, and the independent variable was the place of the consonant. An arcsine transformation was applied to MI, since it is strictly positive. Contrast analyses were performed in all the tests, with /p/ serving as the baseline for L, T2, and, T3, /t/ serving as the baseline for T1 and J, and /k/ serving as the baseline for T4. For the vertical and horizontal components of L, there were significant effects of place, with both /t/ and /k/ showing significantly higher MI than /p/ ($p < 0.01$). For the vertical and horizontal component of T1, there were significant effects of place, with both /p/ and /k/ showing significantly higher MI than /t/ ($p < 0.01$). For the vertical components of T4, /t/ and /p/ had significantly higher MI than /k/. For the horizontal component of T4, /p/ had significantly higher values than /k/, but /t/ was not significantly higher than /k/. For T2 and T3, /p/ had significantly higher MI than /t/ and /k/ in both vertical and horizontal component. For J, there were no significant differences among the places.

B. German coronal data: Manner of articulation effects on MI

The next analysis tested how MI differs by manner of articulation in six coronal consonants, differing in manner of articulation (sibilant, stop, liquid, and nasal). Figure 3(a) shows MI for the vertical dimension for German coronals /t,s,ʃ/, and Fig. 3(b) shows the corresponding data for /d,l,n/. Variability is across the five subjects. Qualitatively, it can be seen that, in the vertical components, the tongue back pellets have higher MI than the jaw and T1, the sonorants have higher MI than the non-sonorants for the jaw, and all of the coronals have roughly the same low MI for T1. It can also be seen that /ʃ/ has lower MI than the rest of the coronals for T3 and T4.

A test on the vertical position of the jaw, with /ʃ/ as the baseline, showed that all /d,n,l/ had higher MI ($p < 0.01$), but there was no significant difference between /s/, /ʃ/, and /t/. An additional test collapsed all the segments, and compared MI of T1y vs T3y. A significantly higher MI for T3y than T1y was found. In addition, a test was performed to test MI for T4 and T3, with /ʃ/ as the baseline. All other coronals showed significantly higher MI than /ʃ/ for T4y ($p < 0.05$), whereas for T3, only the sonorants showed significantly higher MI than /ʃ/. Figure 4 shows MI for the horizontal component of the articulators. There are no qualitative differences between the consonants, and none of the statistical tests taking Articulator or Segment as the independent variable yielded any significant results. As can be seen in Fig. 4, the variability in the horizontal component for all the articulators is quite large, which is expected, since it is the vertical components that are crucial for the constriction achievement.

C. Catalan data: Nasals

Catalan has a contrast between an alveolopalatal nasal /ɲ/ and an alveolar nasal. As is well known palatals highly constrain the tongue body and blade, relative to coronals and dorsals. Therefore this data allows us to determine if the difference between palatals and coronals is quantified by MI.

Figure 5 shows MI for Catalan consonants /n, ɲ/ predicted from V1. Since there are only three subjects, each is marked by a different symbol. There was no attempt to perform statistics due to the small number of subjects, so only qualitative results are presented. We assume a pattern is qualitatively present only when all three subjects exhibit it. This criterion was met for the horizontal components of all three tongue sensors, and the vertical components of the

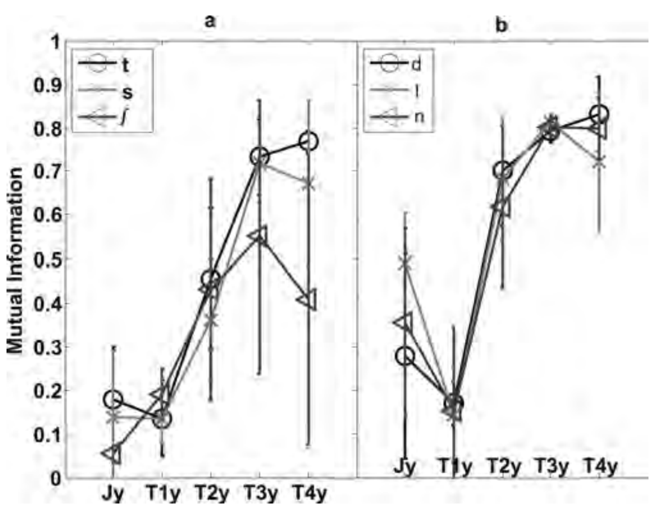

FIG. 3. Mutual information as a function of vertical position of articulators for six German coronals for five subjects. The error bars are centered at the mean for each segment and span a standard deviation on each side of the mean. The data are divided into voiceless consonants /t/, /s/, /ʃ/ (a) and voiced consonants /d/, /l/, /n/ (b).

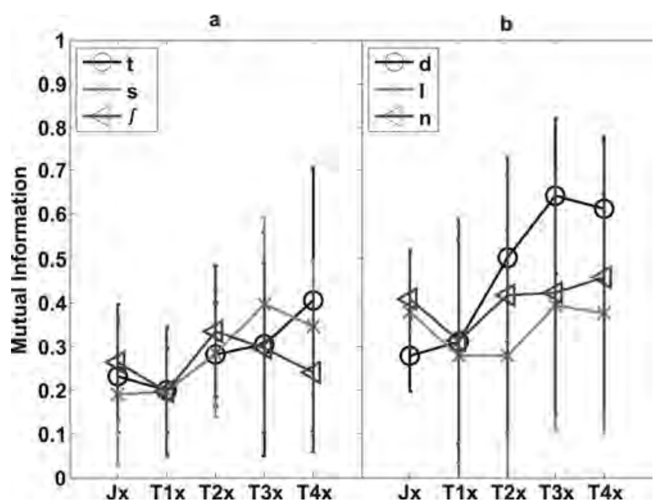

FIG. 4. Mutual information as a function of horizontal component of articulators for six German coronals for five subjects. The error bars are centered at the mean for each segment and span a standard deviation on each side of the mean. The data are divided into voiceless consonants /t/, /s/, /ʃ/ (a) and voiced consonants /d/, /l/, /n/ (b).

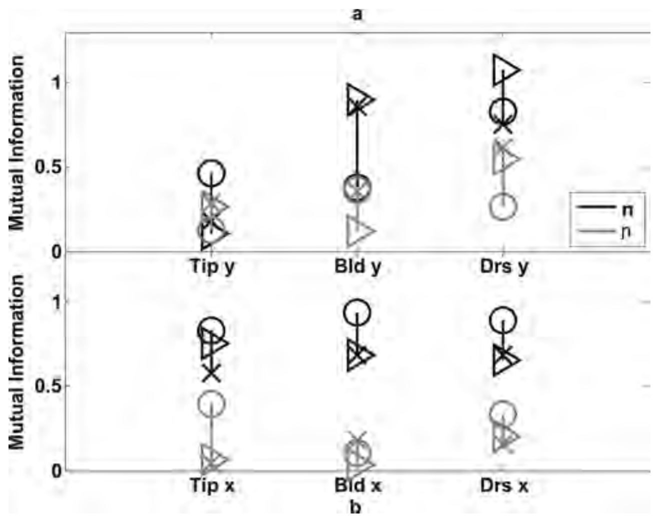

FIG. 5. Mutual information based on V1 as a function of articulator for two Catalan consonants alveolar /n/ and /ɲ/ for three subjects. Labels are as in the original report: tongue sensors for tip, blade and dorsum. Data are divided into vertical (a) and horizontal (b) components. The error bars are centered at the mean value for that consonant and span the range of the data.

tongue dorsum sensor, with MI being lower for the alveopalatal nasal.

D. American English data: Time-course of coarticulation

Figure 6 shows MI as a function of time for the positions of the lower lip pellet of /s/ for 24 subjects of American English, based on production of /sVd/ words. It can be seen that the horizontal component of the lip pellets has higher MI than the vertical component, which was significant ($p < 0.05$). Within the horizontal component, we examined the effect of time on MI. For all times beyond the second, there was a significant difference between the MI at that time and MI at the beginning of /s/ ($p < 0.01$). For the vertical component, however, the only significant differences due to time were between MI at frames nine and 10 ($p < 0.05$) and MI at frame one. Figure 7 shows the temporal effects on MI for the vertical components of the jaw and tongue tip. For both articulators, the only significant effects of time on MI are for the last two time frames. Both exhibit significantly higher MI at frames nine and 10 than at frame one.

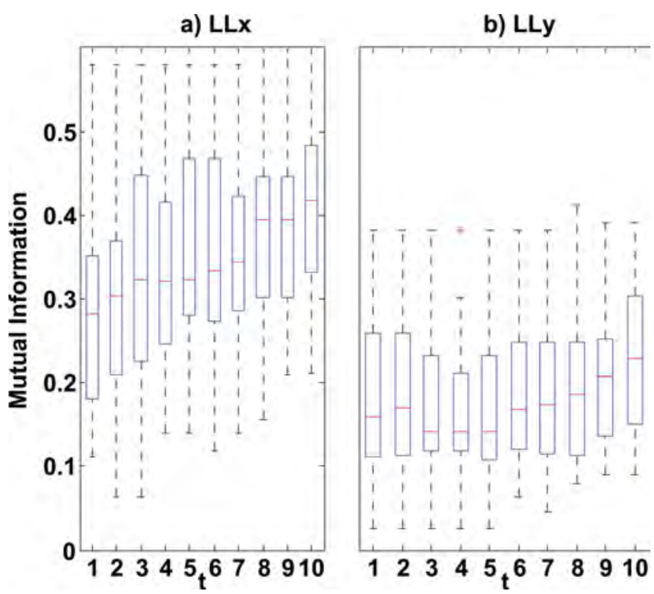

FIG. 6. (Color online) Mutual information as a function of time for the vertical (a) and horizontal (b) components of the lower lip pellet for XRMB data from American English. Variability is across subjects.

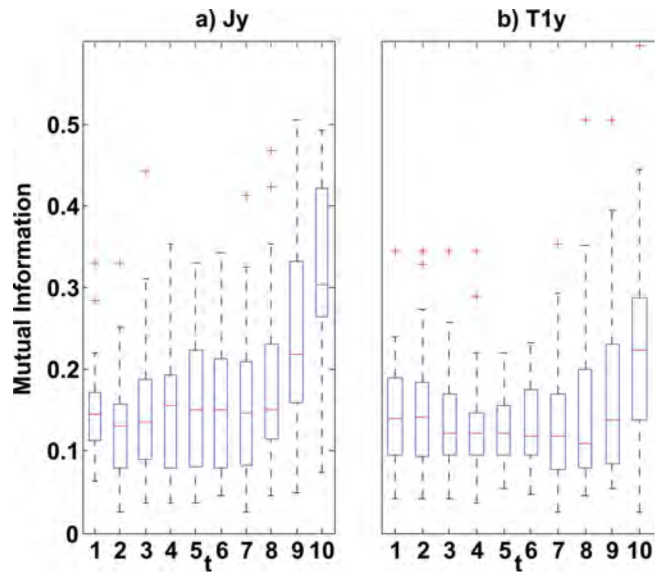

FIG. 7. (Color online) Mutual information as a function of time for the vertical component of the jaw (a) and tongue tip (b) pellets for XRMB data from American English. Variability is across subjects.

E. Locus equations and linearity

Figure 8(a) shows scatterplots of T3y of the consonant /p/ as a function of T3y of the following vowel from the German dataset analyzed in Sec. III, and in Fig. 8(b) the corresponding data for lower lip. Data is shown for all subjects, each with their own symbol. These scatterplots are examples of the functional relations analyzed in locus equations experiments. What we see from these data is that for T3y, for each subject, the dorsum's vertical position in the middle of /p/ is a noisy linear

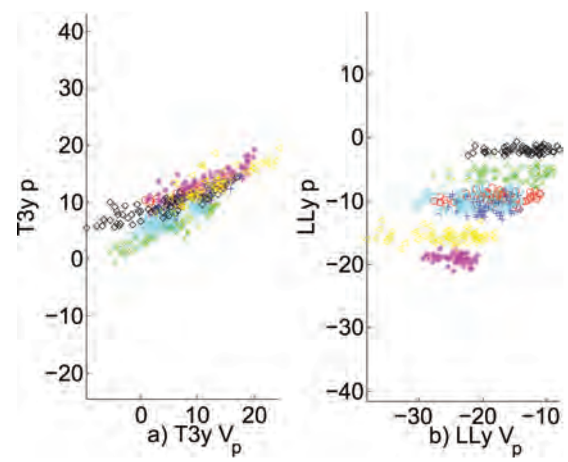

FIG. 8. Scatterplots of the vertical position of the tongue dorsum in the consonant as a function of that for the vowel (a) and for the vertical position of the lower lip (b) in the consonant as a function of the position in the vowel for the German data analyzed in Sec. III. Each color represents a different participant.

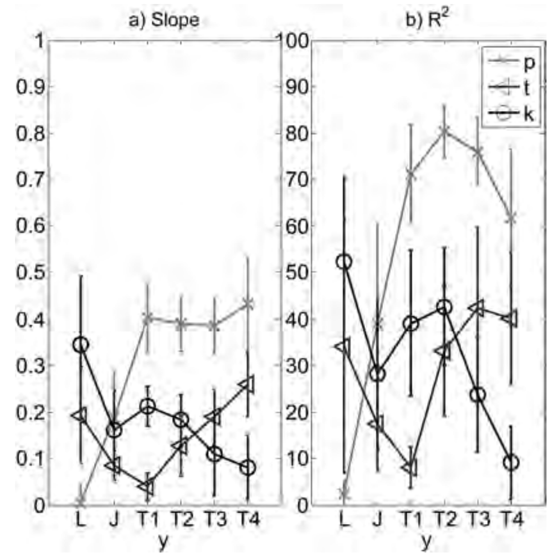

FIG. 9. Slopes (a) and explained variability R^2 (b) for German data analyzed in Sec. III.

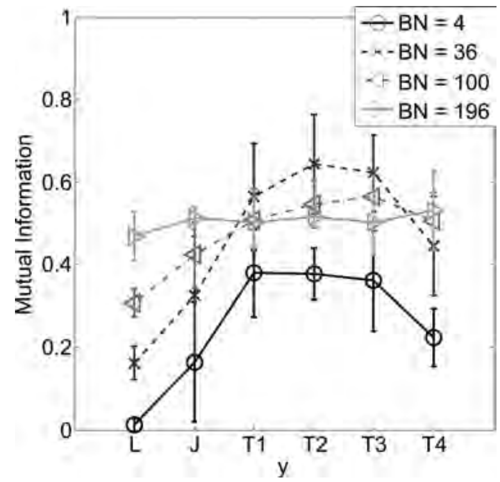

FIG. 10. MI as a function of bin width for the German /p/ data.

function of the vertical position of the dorsum at the middle of the vowel. The same is true for LL, except that the latter's function is quite flat, and hardly any of the variability of the /p/'s lower lip position is explained by the vowel. Figure 9 shows the slopes of regression lines (left) and Explained Variability (R^2) (right) for the *y*-coordinate of the first German dataset analyzed. The figure can be compared directly to that in Fig. 2(a). The patterns of both locus equation slopes and explained variability are very similar to the patterns discussed in Sec. III, and similar in overall pattern to each other. When MI is large, both the slope and R^2 are high, and the reverse.

F. Bin width sensitivity

To measure MI, an estimate of the probability distribution function (pdf) from finite data is necessary. There are many ways of estimating pdfs from finite data. We use the simplest of such estimators, the histogram, in order to make the fewest assumptions about the data, and to use as few parameters as possible. However, there is still one important parameter used in making a histogram estimate of a pdf: histogram bin width. The bin width determines the number of bins into which the data are distributed. The sensitivity of the pdf estimate to bin width depends on how many datapoints are available and how many bins there are. There must be a sufficiently large number of bins to make fine distinctions in pdf shape. But if the number of bins is large, and the amount of data is small, then the occurrence of few data in a bin could be due either to the true pdf having little data in that level of the dependent variable, or it could be due to there not being sufficient data to populate the bins. To measure MI, it is necessary to estimate the two-dimensional joint probability distribution. Even though it is possible to use a different number of bins for the two variables, we always use the same number. To estimate the sensitivity of our estimate of MI to number of bins, we measured MI for the vertical position of the pellets for the /p/ data from Sec. III, while varying the number of bins in each of the two dimensions from two to 14 in steps of four. There were 75 data points for each of the seven subjects in the data set. Figure 10 shows the MI patterns for the vertical dimension of /p/ as the number of bins varies. Throughout this paper, we have used nine two-dimensional bins. Comparison of the patterns seen in Fig. 10 to the data in Fig. 2(a) reveals that the same basic pattern is seen if the number of bins is 4, 9, or 36, which are all smaller than the number of available points. However when the number of bins increases beyond the number of data points available, the basic pattern disappears, as we see for 100 and 196 bins. Moreover, even though the patterns are quite similar for 4, 9, and 36 bins, the overall MI level and details in the shape of the pattern do vary. And it is to be noted that the basic pattern of MI for 4, 9, and 36 bins is quite similar to the pattern for both slope and R2 seen in Fig. 9 for /p/.

V. DISCUSSION

A. MI as a coarticulation/invariance scale

Our German, Catalan, and American English results have shown that MI is an abstract scale on which invariance and coarticulation can be measured. Low MI for a consonant on a specific component of an articulator indicates that the consonant is independent of its vocalic context in determining the position of the articulator in that dimension, and is therefore an indication of an invariant aspect of the consonant's production. High MI, on the other hand, indicates that the consonant is highly dependent on its context for determining the articulators position in a particular dimension, an indication of the presence of coarticulation.

For the German stops, place of articulation influenced the pattern of MI across consonants and articulators. In the labial stop, in which the labial target is specified by the consonant itself, and not by the surrounding vowels, we see very low MI, indicating that the labial target is invariant for labials, as we would expect. On the other hand, the position of the tongue is mostly contextually determined for labials, which is reflected in the high value of MI. This can be directly contrasted with /t/ and /k/, which receive a higher

MI score on the lips, which is expected since lip position is contextually determined for them. But /t/ receives a low MI for the vertical component of the tongue tip and jaw, which accomplish its expected target, and /k/ receives a low MI for the vertical component of the tongue dorsum, which is key in accomplishing its articulatory target.

For the German coronals, the sibilants and the voiceless stop show lower MI for the vertical component of the jaw than the other coronals, indicating higher coarticulation resistance, as also detected in a previous study of this data (Mooshammer *et al.*, 2006). That study also showed that all the coronals exhibit a target for the tongue tip, as expected, as measured by the low MI for all of them, especially as compared with MI for the tongue dorsum for those same consonants.

The analysis of the Catalan data shows that the alveopalatal nasal has greater coarticulation resistance than the alveolar nasal. This can be seen in the lower MI for the alveopalatal nasal, especially in the horizontal position, in agreement with Recasens and Espinosa (2009).

The American English data shows that it is possible to use MI to understand the time-course of coarticulation. The horizontal component of the lip shows higher MI much earlier than the vertical component. We believe that this reflects the greater resistance of /s/ to perturbation of aperture vs its protrusion. It can also be seen that for both jaw and tongue tip, following context has an influence only at the very end of the sibilant, as indicated in the late rise in MI. This is consistent with the Iskarous *et al.* (2011) study of the same data, which comes up with the same conclusion after studying the influence of each vowel type separately.

What emerges from these studies therefore is that MI quantifies invariance and coarticulation in speech production through an identification between resistance and variable independence on the one hand, and coarticulation and variable dependence on the other. The more necessary an articulator's motion for the accomplishment of an action for a segment, the less dependent it is on surrounding segments for its position, and therefore the more independent of context it is. The MI measure may seem unnecessary, however, since the conclusions we can draw from its magnitude have already been drawn from previous empirical studies. However, the MI measure is indeed beneficial, since it allows for quantitative inferences about invariance and coarticulation. It is also valuable since it places invariance and resistance on one end of the scale, identifying them with each other, and coarticulation on the other, establishing the quantitative measure of an important distinction in the study of speech.

The results we obtained on consonants with low MI are very similar to those found for *critical articulators* by the Jackson-Singampalli Statistical Identification model (Jackson and Singampalli, 2009). It is not possible to quantitatively compare their results with ours, since the datasets we used do not provide the extensive sentential data required for their approach, in order for the grand distributions to be representative of as much cross-segmental variability as possible. In future work, we plan to choose a data set that allows us to compare the two methods directly, so that we can determine if they truly yield the same results.

This paper used data from EMA to measure MI. We believe, however, that it is possible to use other forms of quantification of articulatory data for this purpose. Noiray *et al.* (2013) used the LE method to quantify coarticulation of the back of the tongue from ultrasound data, by using the horizontal component of the highest point of the tongue in the ultrasound edge. MI could have also been used in that study and could be based on different types of quantification of whole-tongue images from MRI or ultrasound.

B. Limitations of MI

The current study used MI to quantify invariance and coarticulation by using datasets that have been previously analyzed in the literature to establish if the generalizations previously extracted match those quantified by MI. Within each of the datasets, the generalizations that emerge are quite consistent with those expected and observed elsewhere in the literature. The comparability of MI across the studies is an important issue, since each study has a different number of subjects/tokens and uses different vowels from which to estimate MI. Also, the number and placement of pellets differed across studies, which might also affect the measurement of MI. There are two comparisons that can allow us to address this issue. The two German datasets have the consonant /t/ in them, allowing us to compare MI for the same language. Also, the second German dataset and the Catalan dataset share the consonant /n/. The pattern of MI for /t/ is roughly the same: T1y has lower MI in both datasets than the other vertical components of the tongue, and the horizontal components are roughly the same in MI. What is drastically different across the datasets, however, is that MI in the German stop place of articulation dataset is about half the magnitude in the German manner dataset. There are several reasons for this discrepancy: (1) there are more vowels in the first set, (2) those vowels are both tense and lax while they were only tense in the other; (3) each MI is based on 74 tokens in the first set, but 30 in the second set; (4) the receivers may have been placed at somewhat different locations in the two sets. This study does not allow us to determine which factors, or combination thereof, is responsible for the discrepancy. It is clear, as was seen in Fig. 10, that number of bins, in relation to the number of tokens, is sufficient to affect the level of MI. However, we believe that the number of vowels used is also a major factor, since the entropy of each of the variables (C and V) affects MI. When V includes many vowels in one case and just a few in another, there is likely to be a major effect on MI, since MI is a function of the entire distribution. For a dataset with a few vowels, there are likely to be a few isolated clumps of data, rather than a more spread-out distribution, when there are many vowels, thus affecting the layout of the distribution, and hence MI. For the Catalan/German /n/ comparison, the pattern for the vertical component of the tongue tip and dorsum are the same in both languages, but the MI of the tongue blade is lower for the Catalan data for two subjects. There is potentially a difference in the production of /n/ in the two languages. Again it is not possible for us to determine in this study the source of the discrepancy.

Other studies of MI have dealt with these issues by normalization. Since Shannon proposed MI as a measure of variable independence in 1948, there have been many proposals to normalize MI with different normalization factors and to different ranges, especially to normalize it with respect to the entropy of one of the variables. We have chosen not to use any normalization factors in this study, since the goal was to determine, in general, if MI is reasonable as a measure of the scale from invariance to coarticulation, rather than to study the properties of different normalization factors. In addition, many methods less naive than the histogram have been proposed for multidimensional probability density function estimation, especially kernel density estimation. These methods allow pdf estimates to be very smooth and are not as sensitive to token number and the range of variability of the data. We leave it for further work to investigate how different normalization methods and pdf estimation can improve the estimates and make them more uniform across studies, if that is possible. If not, we still believe MI to be highly informative if done solely within a dataset, since the overall patterns are revealing of the invariance-coarticulation scale.

Another apparent limitation of this approach is that most of the literature on coarticulation focuses on the effects of particular environments (e.g., high vowels) on particular segments (e.g., the fricative /s/). Examples of these types of analyses are plentiful: Iskarous *et al.* (2011); Mooshammer *et al.* (2006); Recasens and Espinosa (2009). Since MI is based on all considered environments at the same time, it does not allow for the description of individual environments. However we believe that most investigations of the effects of individual environments are meant to be incremental, aiming to describe the effects of more and more environments, with the goal of describing invariance-coarticulation behavior of the segment in question, which is exactly what MI tries to accomplish in a single analysis. Of course it is possible to obtain information about the relation between individual vowels and the consonant under study by inspecting the part of the joint distribution, as estimated by the histogram, related to that particular C-V combination. And the general method proposed could be extended to provide other statistical information about the distribution that could indicate certain asymmetries/skewness in the distribution, due to differential vowel effects.

C. Theoretical implications

1. DAC

DAC assigns each segment a number indicating its coarticulation resistance, based on experiments probing how different segments perturb that segment. Even though the theory is a general theory of articulatory constraint, it concentrates on the constraints related to the tongue and jaw. MI accomplishes the aim of DAC, with one major difference: Whereas DAC subjectively assigns each segment a number, MI objectively assigns each segment several numbers, one for each articulator and component. This may seem to complicate the assignment of Coarticulation Resistance indicators; however, we believe that if DAC were to consider all articulators, as the theory eventually aims to do, it would end up assigning many numbers, just as MI does, for each segment. For instance currently, DAC assigns /t/ a higher DAC index than /p/, since /p/ does not resist coarticulation of surrounding vowels as much as /t/. MI does the same for the tongue-related MI numbers (albeit assigning a lower number to the more resistant consonant), but also assigns /p/ a low MI for lips indicating that it resists lip perturbation more than /t/. The issue has already been seen to be a problem in the DAC literature, since a consonant like /k/ is highly resistant in the vertical dimension of tongue back positioning, but not in the horizontal (Recasens and Espinosa, 2009). We therefore believe that MI is an objective extension of DAC theory.

2. Locus equations

Figure 9 shows that MI patterns are quite similar to slope and R2 measures. Furthermore as can be seen in Fig. 8, and as Iskarous *et al.* (2010) showed, articulatory LEs are quite similar to acoustic LEs. The similarity between regression analysis and MI should not be surprising, since both are measures of independence. Slope of a regression line is usually thought to indicate a property of a line, and not independence, *per se*, which is indicated more by explained variability R2. However, if the slope is near 1.0, there is clearly a very straightforward dependence between two variables. As the slope nears 0, whether noise is present or not, a change in the independent variable does not cause a change in the dependent variable. It may seem, however, that there is no need for the MI measure, especially since regression analysis can be conventionally performed with far less data than thought to be sufficient for MI analysis. However, we believe that MI is an extension of LE analysis worthy of further study, since regression assumes Gaussian dependent variables and homoscedasticity, which are unlikely to be met; MI analysis is much less dependent on the distribution of the data. For both DAC and LE, we do not consider MI analysis to be a falsification of these theories, but to be a generalization and an extension, showing that these different theories have the same source: the quantification of the invariance/coarticulation scale.

3. Measuring synergy

Task Dynamics is a dynamical theory of speech production (Saltzman and Munhall, 1989), which attempts to model how articulators move during speech, based on contrastive specifications of gestural parameters. One of the core ideas of the theory is that to carry out a speech task, several articulators collaborate with each other to achieve the task. The articulators form a *synergy* for the achievement of the task. That these synergies exist was shown in several early experiments using perturbations of the articulators during the performance of a task, and observing that other articulators in the synergy compensate for the perturbation (Kelso *et al.*, 1984). But even though the articulators in a synergy collaborate, the theory holds that each articulator's contribution to the synergy may be different. For instance, the lower lip contributes more than the jaw for the achievement of the lip aperture task in a /p/. The theoretical parameter responsible for this effect is termed a *weight*, used in the sense of a mass, so that the greater the weight of an articulator, the more virtually massive it is and the less movable it is. So, for example, the jaw has a higher weight than the lower lip for the achievement of the lip aperture task. So far these weights have been gleaned from subjective evaluation of the results of speech production experiments and through back-fitting of the weight parameters based on speech synthesis using the task dynamic simulator TADA (Task Dynamic Toolkit) (Nam *et al.*, 2004). Another measure of synergy has recently been explored in simulating results from babbling, namely, that the distance between the target positions of different segments can be converted directly into a synergy measure (Nam *et al.*, 2013).

We believe that the MI parameters are a measure of these theoretical weights that specify the synergies in speech production. The reason is that the more important an articulator is for the achievement of a consonant, the more likely it is to resist the perturbation of the surrounding consonant, which is exactly what MI measures. Therefore we predict that the lower the MI of an articulator for a particular consonant, the lower is the weight for the articulator in the achievement of the task for the consonant. This link between articulator independence for a consonant, as measured by MI, and the contribution of a weight for a synergy, as measured by the theoretical weights is confirmed by examining the current settings of the weights (TADA manual). For the oral stops, for instance, the lips have the lowest weight for /p/'s lip aperture task, the tongue tip vertical dimension has the lowest weight for the alveolar gesture for /t/, and the vertical component of the dorsum has the lowest weight for the velar gesture of /k/. Other patterns are also as predicted by MI.

We speculate that there are technical and theoretical consequences of this identification. We believe that the identification could improve speech synthesis from the theoretical model, if the MI estimates are substituted for the subjectively fitted weights. Moreover, the theoretical import of the identification is that the empirically based DAC and LE theories are related through the identification posited here to the theory of task dynamics. Therefore approaches to speech research, which have seemed to be alternative views of the process are shown to be highly related allowing for joint development.

4. Segmental waves

Another theory that MI has implications for is a theory of the control of speech production in which each segment has an associated activation wave which represents the strength with which it influences the vocal tract. This theory has been proposed several times in the history of the field (Öhman, 1966; Fowler and Smith, 1986; Joos, 1948, pp. 109–114). These activation waves would gain in strength during the segment and then fall in strength as overlapped segments start to influence the vocal tract more. However it has always been difficult to obtain direct theoretical support for this theory, since these activation functions represent the time course of strength of the segment, regardless of its context, whereas most empirical studies of coarticulation study individual segments in the context of individual other segments. We believe the time course of MI, as seen in Fig. 6 and Fig. 7, can be regarded as articulator-specific instantiations of the inverse of the segmental strength functions. MI is low at points in the segment, where strength is highest. Even though MI is calculated by taking into consideration the possible contexts of a segment, once it is calculated it represents something about the segment regardless of context, as the strength wave should. Whether the strength wave is articulator-independent or not is a separate issue. But we believe that knowing articulator-based strength waves is crucial to generating an articulator-independent segment wave, if the latter proves to be necessary. MI therefore connects theories of speech that postulate a segmental wave with LE, DAC, and Motor synergy theories that would be suspected to be related to each other, but have remained at an arm's distance from each other.

VI. CONCLUSION

We have shown in this paper that mutual information can be used as a general measure for characterizing the invariance-coarticulation scale. This measure captures known generalizations about speech production, and relates a set of hitherto unconnected theories of the speech process. In future work, we intend to improve the MI measure through appropriate normalization, and to use this measure to probe consonant-to-vowel coarticulation, prosodic effects in speech, and individual differences in coarticulation.

ACKNOWLEDGMENTS

This work was supported by NSF 1246750 to the University of Southern California, NIH-NIDCD grants DC-002717 and DC-008780 to Haskins Laboratories, German Research Council projects TI 69/29 and TI69/32 to Hans Tillmann. We thank Louis Goldstein, Carol Fowler, and Mark Tiede.

Baayen, H. (**2007**). *Analyzing Linguistic Data: A Practical Introduction to Statistics* (Cambridge University Press, Cambridge), pp. 241–300.

Bladon, R. A. W., and Al-Bamerni, A. (**1976**). "Coarticulation resistance in English /l/," J. Phonetics **4**, 137–150.

Brancazio, L., and Fowler, C. (**1998**). "On the relevance of locus equations for production and perception of stop consonants," Percept. Psychophys. **60**, 24–50.

Fowler, C. (**1994**). "Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation," Percept. Psychophys. **55**, 597–610.

Fowler, C. (**1998**). "Coarticulation and extrinsic theories of timing," J. Phonetics **8**, 113–133.

Fowler, C. A., and Smith, M. R. (**1986**). "Speech perception as 'Vector Analysis': An approach to the problems of invariance and segmentation," in *Invariance and Variability in Speech Processes*, edited by J. Perkell and D. Klatt (Erlbaum, Hillsdale, NJ), pp. 123–139.

Hoole, P. (**1999**). "On the lingual organization of the German vowel system," J. Acoust. Soc. Am. **106**, 1020–1032.

Iskarous, K., Fowler, C., and Whalen, D. H. (**2010**). "Locus equations are an acoustic expression of articulator synergy," J. Acoust. Soc. Am. **127**, 2021–2032.

Iskarous, K., Shadle, C. H., and Proctor, M. I. (**2011**). "Articulatory-acoustic kinematics: The production of American English /s/," J. Acoust. Soc. Am. **129**, 944–954.

Jackson, P., and Singampalli, V. D. (**2009**). "Statistical identification of articulation constraints in the production of speech," Speech Commun. **51**, 695–710.

Joos, M. (**1948**). "Acoustic Phonetics," *Language Monographs, No. 2* (suppl. 24) (University of Chicago Press, Chicago, IL), pp. 109–114.

Kelso, J. A. S., Tuller, B., Bateson, E. V., and Fowler, C. A. (**1984**). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures," J. Exp. Psychol.: Hum. Percept. Perform. **10**, 812–832.

Krull, D. (**1987**). "Second formant locus patterns as a measure of consonant-vowel coarticulation," PERILUS **5**, 43–61.

Krull, D. (**1989**). "Consonant-vowel coarticulation in spontaneous speech and reference words," PERILUS (Institute of Linguistics, Stockholm University) **10**, 101–105.

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

Lindblom, B. (**1963**). "Spectrographic study of vowel reductions," J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B., and Sussman, H. (**2012**). "Dissecting coarticulation: How locus equations happen," J. Phonetics **40**, 1–19.

Mooshammer, C., Hoole, P., and Geumann, A. (**2006**). "Inter-articulator cohesion within coronal consonant production," J. Acoust. Soc. Am. **120**, 1028–1039.

Nam, H., Goldstein, L., Giulivi, S., Levitt, A., and Whalen, D. (**2013**). "Computational simulation of CV combination preferences in babbling," J. Phonetics **41**, 63–77.

Nam, H., Goldstein, L., Saltzman, E., and Byrd, D. (**2004**). "TADA: An enhanced, portable Task Dynamic model in MATLAB," J. Acoust. Soc. Am. **115**, 2430.

Noiray, A., Ménard, L., and Iskarous, K. (**2013**). "The development of motor synergies in children: Ultrasound and acoustic measurements," J. Acoust. Soc. Am. **133**, 444–452.

Öhman, S. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**, 151–168.

Recasens, D. (**1985**). "Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences," Lang. Speech **28**, 97–114.

Recasens, D. A. (**2012**). "A study of jaw coarticulatory resistance and aggressiveness for Catalan consonants and vowels," J. Acoust. Soc. Am. **132**, 412–420.

Recasens, D., and Espinosa, A. (**2009**). "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan," J. Acoust. Soc. Am. **125**, 2288–2298.

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

Shannon, C. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.

Stevens, S., and Blumstein, S. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Sussman, H. M., McCaffrey, H. A., and Mathews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**, 1309–1325.

Syrdal, A. (**1985**). "Aspects of a model of the auditory representation of American English vowels," Speech Commun. **3**, 121–135.

Westbury, J. (**1994**). *X-ray Microbeam Speech Production Database User's Handbook* (University of Wisconsin, Madison, WI), pp. 88–107.