Quantifying lingual coarticulation in German using mutual information: An ultrasound study[a)]

Dzhuma Abakarova,[1,b)] Khalil Iskarous,[2] and Aude Noiray[1]
[1]*Department Linguistik, University of Potsdam, 14476 Potsdam, Germany*
[2]*Department of Linguistics, University of Southern California, Grace Ford Salvatori 301, Los Angeles, California 90089-1693, USA*

(Received 27 May 2017; revised 11 June 2018; accepted 6 July 2018; published online 21 August 2018)

In previous research, mutual information (MI) was employed to quantify the physical information shared between consecutive phonological segments, based on electromagnetic articulography data. In this study, MI is extended to quantifying coarticulatory resistance (CR) versus overlap in German using ultrasound imaging. Two measurements are tested as input to MI: (1) the highest point on the tongue body and (2) the first coefficient of the discrete Fourier transform (DFT) of the whole tongue contour. Both measures are used to examine changes in coarticulation between two time points during the syllable span: the consonant midpoint and the vowel onset. Results corroborate previous findings reporting differences in coarticulatory overlap in German and across languages. Further, results suggest that MI used with the highest point on the tongue body captures distinctions related both to place and manner of articulation, while the first DFT coefficient does not provide any additional information regarding global (whole tongue) as opposed to local (individual articulator) aspects of CR. However, both methods capture temporal distinctions in coarticulatory resistance between the two time points. Results are discussed with respect to the potential of MI measure to provide a way of unifying coarticulation quantification methods across data collection techniques. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5047669

[JFL] Pages: 897–907

I. INTRODUCTION

Coarticulation, the temporal overlap of phonetic segments, is a fundamental characteristic of speech and a major source of articulatory and acoustic variability. A complete description of this variability requires accounting for coarticulation resistance (CR), the degree to which a segment is susceptible to or resists overlap with consecutive segments, due to its inherent articulatory properties (Recasens and Espinosa, 2009; Iskarous *et al.*, 2010). To address the need for a unified approach to quantifying CR, Iskarous *et al.* (2013) recently adapted mutual information (MI), a quantity defined within information theory (Shannon, 1948), to the quantification of CR in three languages: American English, Catalan, and German. Within the MI approach, each segment is assigned a number on a coarticulation/invariance scale based on the amount of physical information (e.g., the position of articulators) that it shares with its neighbors. When two consecutive segments show great coarticulatory overlap, it suggests the two segments share physical properties (e.g., a similar tongue position) and therefore the MI value is high. On the contrary, when the tongue position for a segment conflicts with that adopted by its consecutive segment, the MI value is low. Using articulatory data collected with electromagnetic midsagittal articulography (EMA), Iskarous *et al.* (2013) showed that MI values across consonant-vowel (CV) sequences differ as a function of consonants' place and manner of articulation in three languages: American English, Catalan and German (more details on the measurements of Iskarous *et al.* and findings are provided in Sec. II B).

Capitalizing on the findings from Iskarous *et al.* (2013), this study aims to contribute to establishing a unified measure of CR, by applying the MI method to the quantification of lingual V-to-C coarticulation measured with ultrasound imaging. Ultrasound imaging technique differs from EMA in that it provides a continuous tongue contour (versus fixed point coordinates) and therefore a more detailed representation of the tongue body (see Sec. II C for more details). In this study, we examined the validity of two tongue shape quantification methods as input for MI calculation: the highest point on the tongue body and the first coefficient of the discrete Fourier transform (DFT). These methods allow for comparing a global measure of lingual coarticulation (the whole tongue, here quantified with DFT) with a local measure (individual articulators, here quantified with the highest point on tongue body) used in the original study of Iskarous *et al.* (2013). Iskarous *et al.* (2013) used data from multiple databases. In contrast, the present study reports on data from a single database of German adult speakers to prevent possible methodological confounds due to differences in data collection (e.g., differences in the phonological, phonetic, and prosodic properties of the production material) and in data processing (e.g., heterogeneity in phonetic labeling and tongue detection procedures).

Before describing the specifics of the MI method, Sec. I A provides an overview of two very common methods

[a)]Portions of this work were presented at *Labphon 15*, Ithaca, NY, USA, 2016 and *Abstraction, Diversity, and Speech Dynamics Workshop*, Munich, Germany, 2017.
[b)]Electronic mail: dzhuma.abakarova@uni-potsdam.de

employed to describe coarticulatory overlap versus resistance in adults' speech, namely, locus equations (LEs) (Lindblom, 1963) and degree of articulatory constraint (DAC) model (Recasens *et al.*, 1997).

A. LE and DAC models

The LE approach has been widely employed in speech production research to numerically characterize coarticulation degree in CV sequences where C is an oral stop. The slope of the regression line relating F2 at the onset of a given vowel to F2 at the midpoint of a variety of vowels (Lindblom, 1963) quantifies the magnitude of acoustic change in the domain of a given consonant as the vowel varies (Krull, 1987; Fowler and Brancazio, 2000). Hence, the magnitude of the regression slope has been evidenced to reflect consonants' place of articulation. For instance, CV sequences involving labial stops exhibit greater coarticulatory overlap and therefore are characterized by steeper slopes than sequences including alveolar stops (e.g., Lindblom and Sussman, 2012; Sussman *et al.*, 1991; for a review, see Iskarous *et al.*, 2010). The studies that investigated the articulatory mechanisms underlying the systematic differences in LEs' intercepts and slopes suggest that regression intercepts directly reflect the degree to which the tongue body (estimated via F2 in the acoustic speech signal) contributes to the formation of a lingual consonantal constriction, while the slopes reflect consonants' resistance to coarticulation associated with the horizontal motion of the tongue back (e.g., Iskarous *et al.*, 2010). The linear relationship produced by the y-intercept of the LE plotted against its slope and referred to as second-order locus equation (SOLE) (Chennoukh *et al.*, 1997) is also likely to emerge from the relation between the horizontal position of the tongue back at the onset of the vowel and that its midpoint (e.g., Iskarous *et al.*, 2010). However, the traditional LE approach may not capture subtle differences between consonants that have the same place of articulation (e.g., differences in manner of articulation, Tabain, 2000). Moreover, LEs were originally intended for studying CV sequences, and extending the approach to other types of sequences (e.g., V-V) may require model modification (e.g., Lindblom and Sussman, 2012).

Expanding on the original acoustic LE, a number of articulatory studies (e.g., Farnetani and Recasens, 1993; Iskarous *et al.*, 2010; Noiray *et al.*, 2013) have employed linear regressions to see how well the constriction location for a given segment is predicted by that of its neighbors. One major advantage of an articulatory approach to quantifying CR is the possibility of disentangling the horizontal and vertical components of coarticulatory resistance. Further, the linear regression analysis of articulatory data is not restricted to reflecting the tongue body's position for one segment with respect to another; the method can be applied to describing properties of other articulators (e.g., the lips and the jaw) as well as manner related differences in articulation (Iskarous *et al.*, 2013).

However, employing a linear regression to gain insight into the articulatory mechanisms responsible for variation in coarticulation may presents some limitations. The main one regards the strong but not necessarily valid assumptions made about the data distribution. For instance, a linear regression assumes that the relationship between an articulator's position at two different time points is linear, but that may not always be the case (e.g., in vowel-to-vowel coarticulation or in F2 dynamics for vowels: Sóskuthy, 2017). Even when the assumptions of a linear model are met, there remains the issue of sufficient descriptiveness. Since regression coefficients are estimated based on the first and second order statistics in a linear way, they can reveal monotonic but not higher order dependencies. The interpretation of the results may be problematic at times. Specifically, the slope of a regression line usually indicates a property of a line, whereas the independence or the strength of the relationship is indicated by explained variability (R^2). However, in studies with repeated measurements for which mixed models are more suitable to account for the non-independence of errors, obtaining a correlation coefficient becomes a non-trivial problem with no generally accepted solution so far (e.g., Roberts *et al.*, 2011).

The DAC model of coarticulation developed by Recasens *et al.* (1997) assesses coarticulatory resistance based on the DAC associated with a particular segment on the tongue. The model assigns each speech segment a DAC value based on the average Euclidean distance between the position of articulators during consonant production in different vocalic contexts and its centroid across vocalic contexts. In this approach, segments are categorized into three groups representing different levels of resistance to coarticulation. For instance, consonants with the lowest resistance belong to the first group (e.g., bilabials, schwa); the ones with the highest resistance are grouped in the third group (e.g., palatals, trills). Despite providing a method for categorizing consonants as a function of their contextual resistance, the DAC scale remains a qualitative rather than quantitative description of segments' resistance.

B. MI measure of coarticulation resistance

Iskarous *et al.* (2013) adapted the measure of MI to assess CR by combining the insights of the LE (Lindblom, 1963) and the DAC model of coarticulation (Recasens *et al.*, 1997) approaches while aiming to address their limitations. MI measures the general (linear and nonlinear) dependence between two random variables (Smith, 2015). It indicates the amount of information on random variable X obtained by observing random variable Y. MI is one of the basic notions of information theory (Shannon, 1948) and it is strongly related to the measure of information, entropy. Entropy (H) is a measure of uncertainty about a random variable that is associated with the probability of an event.

To illustrate these notions in terms of coarticulation, let us take as an example the production of a CV syllable. In this case, C represents the random variable X while V is the random variable Y. When examining changes in tongue position over time during the production of a CV syllable, we first observe the tongue position at the acoustic midpoint of C (variable X). One may ask how much information about the tongue position for the upcoming V (variable Y) is available given the observation of the tongue position for C (variable X) at its acoustic midpoint. In this case, *information* refers to the knowledge about the position of the tongue. If the tongue shape associated with X does not provide observers with any information about the tongue shape associated with the upcoming segment Y, all possible outcomes in the domain of Y remain equally probable; Y can be any segment in a given language. If possible, Ys have different probabilities of occurrence after observation of X, then the uncertainty about Y is decreased due to the observation of X. In other words, H(Y|X) is lower than H(Y). This is the case when coarticulatory effects are present: in case of a syllable such as /bi/ by observing tongue raising during the production of /b/, observers can at least predict that the following vowel is likely to be high; thus, the uncertainty about variable Y is reduced. MI(X;Y) represents the reduction in uncertainty about variable Y as a result of observing variable X. If the MI between two variables is high, then the entropy of variable Y is reduced. If MI between two variables is zero, the two variables are independent of each other, and observing one does not reduce uncertainty about the other. Note that MI cannot be lower than zero because the observation of variable X may or may not provide some additional information about variable Y but it cannot reduce the information about Y that was already available. Figure 1 illustrates the relationship between the marginal entropies of two random variables, H(X) and H(Y), their respective conditional entropies, H(X|Y) and H(Y|X), and their MI, MI(X;Y), as well as the relationship between MI and the scale between invariance and coarticulation scale. Here, if the MI value is low on the scale from 0 to 1, the segments X and Y do not share much information and thus can be described as invariant, or independent of each other. Conversely, an MI value closer to 1 indicates the presence of large coarticulatory effects between the two segments.

MI is conceptually close to linear regressions (used, for instance, in the original LE studies) in that both methods measure general interdependencies between variables. However, unlike linear regressions, MI is a nonparametric measure that does not have to rely on any inherent assumptions about the data distribution and can thus take into account both linear and non-linear dependencies (e.g., Darbellay, 1999). Such a non-parametric approach allows for unbiased, data-driven investigations of complex patterns.

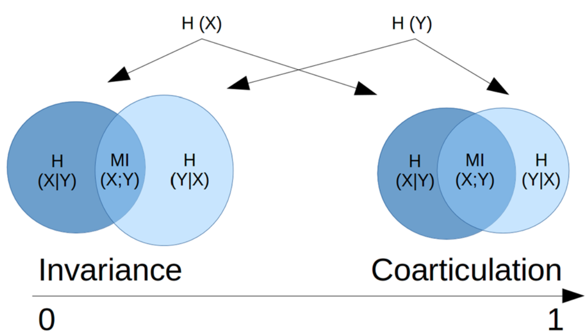

FIG. 1. (Color online) Schematic representation of the relationship between the entropy of two random variables, H(X) and H(Y), respectively, MI(X;Y), and coarticulation-invariance scale.

Hence, it has become widely used in statistics, machine learning, and computational neuroscience (e.g., Gao *et al.*, 2015).

Iskarous *et al.* (2013) used the MI method to examine the dependency between EMA measures taken within segments in four spoken language corpora from three languages (English, German, and Spanish). In this framework, high MI values were found for CV sequences exhibiting large coarticulatory overlap while low MI values were observed when measures taken from neighboring segments were relatively independent of each other. For German consonants, MI differed across both places and manners of articulation. The comparison of the MI method with an articulatory adaptation of standard LE analysis (as in Noiray *et al.*, 2013) showed that both slopes and correlation coefficients assessing the strength of the linear relationship between data at two time points (R^2) are high when MI is high.

C. MI measure applied to ultrasound imaging

Iskarous *et al.* (2013) have suggested that MI can be applied to various forms of articulatory data quantification obtained with techniques other than EMA, for example, ultrasound imaging or MRI. In this study, we use the MI statistic to calculate CR for German consonants based on ultrasound data. The many advantages of ultrasound scanners (e.g., mobility, affordability, and non-invasiveness) have made it an increasingly popular tool in the field of speech production. Ultrasound is currently the most fruitful way to obtain quantitative lingual data from young children (e.g., Noiray *et al.*, 2018; Noiray *et al.*, 2013), sensitive populations (e.g., in participants with developmental apraxia of speech: Nijland *et al.*, 2002), and in speakers living in remote areas: Whalen *et al.*, 2011). However, ultrasound data present methodological challenges primarily due to the absence of physiological reference points.

Contrary to EMA data that represent tongue position in terms of xy coordinates of fixed points on the tongue, ultrasound data consist in tongue contours displayed in the midsagittal or coronal planes that can be quantified in various ways. Some research has focused on quantifying the position of specific tongue articulators (tip, body, back) (e.g., Noiray *et al.*, 2014; Noiray *et al.*, 2013; Recasens and Rodríguez, 2017), while other methods preferred characterizing the whole tongue shape (e.g., Dawson *et al.*, 2015; Scobbie *et al.*, 2013; Zharkova *et al.*, 2015). The output of many of these methods can be used in MI analyses and thus enable comparisons between *local* (individual articulators) and *global* (the whole tongue) measures of CR. While local measures of CR are important for understanding the relative role of different articulators in the production of segments, global measures of CR that describe the whole tongue configuration allow for probing effects of voicing and manner as well as effects of prosodic factors such as stress, speech rate, or syllable position.

In the present study, we compared two approaches employed for quantifying lingual coarticulation between adjacent segments: the highest point of the tongue body and the whole tongue contour as described by the first coefficient of the DFT. The highest point on the tongue body is used here as an absolute measure of tongue height and frontness (e.g., Noiray *et al.*, 2013), and it does not correspond to a measure of tongue height relative to the palate (e.g., Ménard *et al.*, 2012). The second method focuses on the whole tongue contour as described by the first coefficient of the DFT that captures the gross features of the tongue shape (Liljencrants, 1971). A detailed description of the procedure is outlined in Sec. II. In the present study, this method is used to examine to what extent the tongue body's resistance to coarticulatory overlap with adjacent vowels reflects the resistance of the whole tongue. The highest point of the tongue metric therefore provides a local measure of coarticulatory resistance while the DFT output provides a more global measure of CR. Both methods were employed at two time points within a CV syllable, the consonant midpoint and offset of the consonant (or vowel onset), to examine the temporal organization of coarticulatory resistance.

To estimate CR differences related to consonants' place of articulation, we compare MI values for German stops /b/, /d/, /ɡ/. We generally expect a pattern similar to that observed by Iskarous *et al.* (2013) for the primary articulators in German place of articulation data. Regarding the horizontal position of the highest point on the tongue body which assesses variation in the font-back dimension, we expect the bilabial stop /b/ to have the highest MI, the alveolar stop /d/ to have the lowest MI and the velar stop /ɡ/ to have an intermediate MI. For the vertical position of the tongue that estimates variation in height, we expect the bilabial /b/ to show the highest MI values and the alveolar /d/ to have the lowest MI. However, the velar stop /ɡ/ is more likely to pattern similarly to the alveolar /d/ in the vertical dimension. While /ɡ/ is associated with rather low CR in the front-back dimension (Öhman, 1966), studies suggest it is more invariant in the vertical dimension than in the horizontal one due to the necessity of making a full constriction on the palate (e.g, review in Fowler and Brancazio, 2000).

The alveolar stop /d/ is compared to the alveolar fricative /z/ in order to examine whether the effect of manner of articulation on coarticulation patterns is captured by the MI values. As for the alveolar fricative /z/, we expect the MI value for the horizontal dimension to be approximately the same as for the alveolar stop /d/, because they share the same place of articulation and thus constriction location. However, we expect the alveolar fricative /z/ to be less variable, i.e., have a lower MI value, in the vertical dimension than /d/ because of the difference in constriction degree and the fact that the positioning of the tongue body should be more precise for /z/ than for /d/. We predict that the first DFT coefficient will generally have a similar pattern to that shown by the horizontal position of the highest point of the tongue body. However, the DFT may show a higher MI for /z/ than for /d/, because it is not as constrained at the back of the tongue as it is in the case of /d/ (Recasens and Rodriguez, 2017).

Finally, as regards CR differences between the two time points within the consonant, we expect information sharing between the vowel onset and the vowel midpoint to be greater than that between the consonant midpoint and the vowel midpoint. According to the concept of a gestural activation wave (e.g., Fowler and Brancazio, 2000; Fowler and Saltzman, 1993), the gestures for consonants or vowels first strengthen, then weaken over time. Thus, the consonant's influence over vocal tract shape would be strongest during the consonant closure and become weaker by the vowel onset, giving the vowel gesture the opportunity to contribute more to the tongue positioning.

II. METHOD

A. Participants

Participants were 12 (6 females) native speakers of German (mostly students at the University of Potsdam, Brandenburg), aged between 19 and 28 years [$M = 23.2$, standard deviation (SD) $= 2.9$], with no reported history of speech, language or hearing difficulties. They were all recorded at the Laboratory for Oral Language Acquisition at Potsdam University.

B. Stimuli

The stimuli were disyllabic /C1VC2ə/ non-words, respecting German phonotactics. The first, target syllable consisted of one of the four consonants /b/, /d/, /ɡ/, or /z/ and one of the six long vowels /i/, /y/, /u/, /a/, /e/, or /o/. The second syllable also contained one of the consonants above and the vowel schwa (/ə/). The first and the second syllables were combined in such a way that C1 was never the same as C2. Stress was always on the first syllable. The disyllables were preceded by the carrier word "eine" (/a ɪ n ə/) which ends with the neutral vowel /ə/, selected to approximate a neutral tongue position and thus prevent strong coarticulatory effects on the target syllable.

The set of stimuli was chosen for planned comparison with children. The consonants are voiced because one of the project's aims is to compare acoustic and articulatory measures of coarticulation, and the voiced consonants are easier to measure acoustically; that is, the burst of the consonant is more reliably detectable in voiced consonants. Further, target vowels were chosen to maximally approximate the vowel space, by including the opposition of front and back vowels (e.g., from /i/ to /u/) as well as high and low vowels (e.g., from /i/ to /a/) and round-unrounded vowels (e.g., from /i/ to /u/ or /y/).

The C1Vs syllables examined in this study were designed as a fully crossed set of Cs and Vs to which the second syllable was added, and C2 always differed from C1. It resulted in 24 target syllables (4 C1 × 6 V = 24 C1V) in three consonantal contexts making a total of 72 stimuli. The stimuli were divided into nine blocks containing every C1V pair once and every C1-C2 combination only twice. The order of stimuli within each block and the order of blocks were randomized. The nine blocks, with 24 stimuli each, resulted in 216 productions per participant with nine repetitions of every target syllable and 54 repetitions of every consonant (1 consonant × 6 vowels × 9 blocks). Overall, this resulted into a total of 648 repetitions per consonant across subjects (54 repetitions per participants × 12 participants) and 2592 observation of target consonants across participants. The pseudowords were embedded in a carrier phrase with the German female article /aɪnə/ resulting in short utterances such as /aɪnə biːdə/. Consonant-vowel coarticulation was measured between the full vowel of the pseudowords and the preceding consonant.

C. Procedure

Prior to the recording, participants were familiarized with the production task and experimental procedure. They were comfortably seated in a barber's chair with their chin positioned on the ultrasound probe. They were asked to avoid moving while seated and to look at a fixed visual target located in front of them.

The production task consisted in a repetition of the stimuli described above that were presented auditorily. The auditory stimuli were pre-recorded with an adult female native speaker of German, who was naive to the purpose of the recording. The model speaker, as well as all the participants, spoke Standard German with no audible dialect. During the recording, the auditory stimuli were prompted via speakers located on both sides of the participant using custom-made scripts for MATLAB (version R2016a, The MathWorks, Natick, MA) within the sonographic and optical linguo-labial articulation recording (SOLLAR) platform (Noiray *et al.*, 2015). We chose this elicitation method over a reading task to enable future comparison with children's speech and because reading involves additional grapheme and phonological decoding processes that were not of interest in this study.

Articulatory data were collected with SOLLAR (Noiray *et al.*, 2015). This platform allows for simultaneous recordings of tongue movements via ultrasound imaging (Sonosite Edge, sampling rate: 48 Hz), audio speech signals with a microphone (Sennheiser, sampling rate: 48 kHz) and video data from a camera (SONY HDR-CX740VE, sampling rate: 50 Hz). The probe was mounted on a custom-made spring-loaded probe holder that allows the probe to move vertically following the natural motion of the jaw but prevents motion in the lateral and horizontal planes.

D. Analysis

1. Measurement points

First, the acoustic signal was automatically segmented and labeled using WEB-MAUS (Kisler *et al.*, 2016). The labeling for target segments was then checked and manually corrected in PRAAT, version 6.0.04 (Boersma and Weenink, 2015). The onset of stop consonants was identified as the beginning of the closure phase. For the fricative, the onset and offset were identified at the beginning and the end of frication. Stable periodic cycles in the oscillogram as well as a stable formant pattern, especially a clearly detectable second formant (F2), were used as indices for the onset and offset of vocalic segments. The first ascending zero-crossing in the oscillogram at the beginning of the periodicity was used for the vowel onset, the first ascending zero-crossing after the end of periodicity and disappearance of F2 as the beginning of the following consonant.

From the resulting intervals, the time stamps for the temporal midpoint of the consonant (C50), the temporal onset of the vowel (V00) and the temporal midpoint of the vowel (V50) were used for subsequent manual tongue detection in custom MATLAB scripts within the SOLLAR platform. All tongue contours were checked by a second experimenter and corrected if necessary. Then, the xy coordinates of the highest point of the tongue body (TB) and of the 100 points of the tongue contour were extracted and subsequently used in the MI analysis.

2. Tongue shape quantification

Two methods of tongue shape quantification are used as input for MI calculation. In the first step, MI is calculated for the horizontal and vertical coordinates of the highest point of the tongue body. In the second step, the MI values are obtained for the whole tongue contour, i.e., for the hundred points obtained as an output of the SOLLAR tongue detection procedure. To enable calculation of MI for the whole tongue contour the dimensionality of the latter is reduced by applying the DFT procedure. Tongue contours are transformed into the spatial frequency domain following Dawson *et al.* (2015) adaptation of the method developed by Liljencrant (1971) to transform x-ray images of the tongue. Dawson *et al.* (2015) modified the method for ultrasound data, specifically, for the absence of a static reference point for a coordinate system by transforming the tangent angle values for each point on the tongue contour as a function of arc length. One-dimensional 100-point DFT of the resulting array of angles in radians for each tongue contour is computed with the fast Fourier transform (FFT) as implemented in the "numpy.fft.rfft" function (Oliphant, 2006) in PYTHON (version 2.7.10, Python Software Foundation) with the default normalization. The output of this procedure provides coefficients representing each tongue contour as a sum of sine and cosine waves of increasing frequency. A coefficient can also be represented in terms of the magnitude (length) and phase (direction) that describe its location in radial coordinates. For the purposes of this analysis a contour is represented by "real" (cosine coordinates) and "imaginary" (sine coordinates) parts of the first coefficient of the FFT that corresponds to the largest scale features of the tongue shape.

3. "MI calculation"

To test to what extent the tongue position during a consonant (C) production is predicted from the tongue position during the following vowels (V), we calculate the amount of information shared between the two segments using the method employed by Iskarous *et al.* (2013). In this method, the joint distribution of C and V is compared to the hypothetical joint distribution that is based on the assumption that V and C are independent. MI between the two distributions is calculated according to the following formula:

$$MI(X,Y) = \sum_{x=X} \sum_{y=Y} p(x,y) \log_z \frac{p(x,y)}{p(x)p(y)},$$

where X = {x1, x2,…, xn} denotes the tongue's position during the consonant, and Y = {y1, y2,…, yn} is the tongue's position observed during the vowel; p(x,y) is the probability of the measured joint distribution of x and y, and p(x)p(y) is the joint probability of the distribution of x and y assuming that X and Y are independent.

In this study, MI for each consonant is calculated on a speaker-by-speaker basis by measuring its position as a function of the positions of different vowels. For that, two probability distributions are estimated, joint probability distribution of C and V (where C is always the same and V is all the six vowels), and their independent joint distribution that is based on the assumption that C and V are independent. Both probability distribution functions are estimated using the histogram method. The method works by binning the data into a 5-by-5 grid of equally spaced two-dimensional containers where the two dimensions correspond to the tongue position during each instance of the target consonant and the tongue position during the following vowel. The number of bins is chosen subjectively, taking into account the number of data points available. It is important that the number of bins does not exceed the number of data points so that there are no empty bins. The larger the difference between the joint distribution and the distribution under the assumption of independence, the higher is the MI between two segments. Figure 2 provides an illustration of MI calculation for the labial stop /b/ and the alveolar stop /d/. Here, the joint distribution for /b/ (a) appears quite different from the independent joint distribution (b) whereas for /d/ the two distributions [(c), (d)] do not differ as much. The MI calculation was conducted on (a). The calculation was conducted in R (version 3.4.0, R Foundation for Statistical Computing,Vienna, Austria).

4. Statistical analysis

To test whether the mean by-subject MI values differ significantly between consonants, we fit linear mixed models (LMMs), with MI values as the response variable, consonant as predictor and by-subject random intercepts. MI values were arcsine-transformed because they are strictly positive. Consonant was a four-level (/b, d, g, z/) factor treatment-coded with "b" as baseline. All pairwise comparisons for the four levels of the consonant factor were obtained with the help of the *glht* function from the "MULTCOMP" package, version 1.4-7 (Hothorn *et al.*, 2008) by manually setting the contrast matrix.

To test whether the mean information sharing is significantly different between consonant midpoint and vowel onset for each consonant, we fit LMMs, with MI values as the response variable, condition as predictor and by-subject random intercepts. Condition was an eight level factor combining consonant identity and time point with "b-mid" as baseline. MI values were arcsine-transformed because they were strictly positive. All models were fit using the R package "LME4," version 1.1-14 (Bates *et al.*, 2015). To account for multiple comparisons, the *p*-values were corrected following the truncated closed test procedure suggested in Westfall (1997) as implemented in the "MULTCOMP" package, version 1.4-7 (Hothorn *et al.*, 2008).

III. RESULTS

Figure 3 displays the means and the SDs of the MI values obtained for the twelve German speakers between the

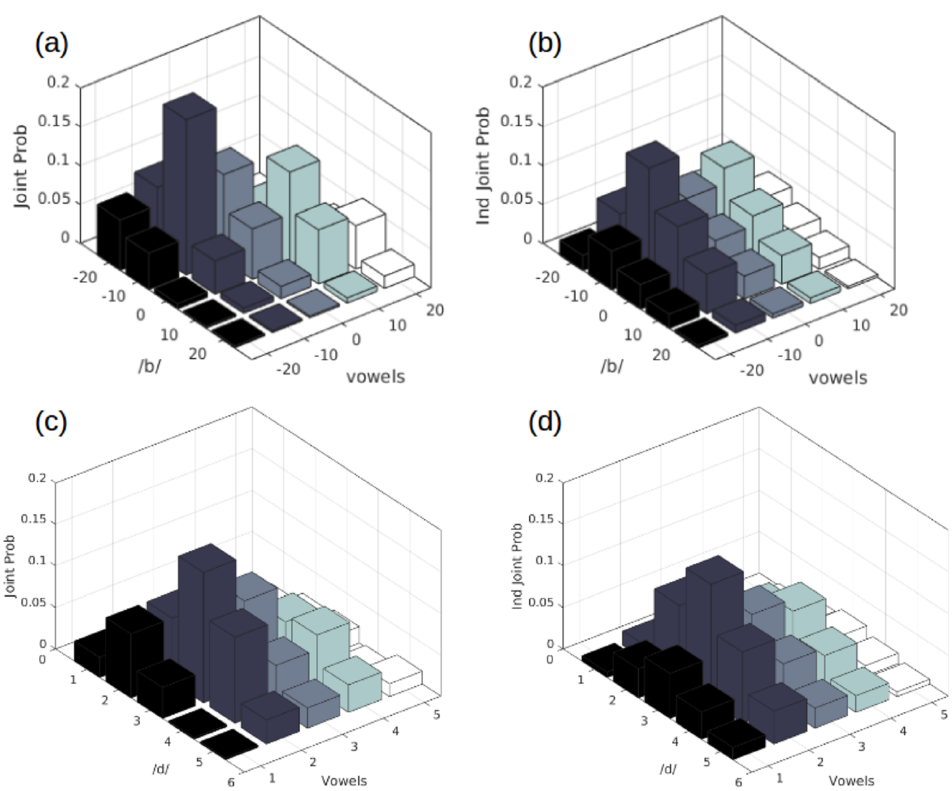

FIG. 2. (Color online) Example of MI calculation for the labial stop /b/ (upper panels) and alveolar stop /d/ (lower panels). Panels (a), (c) show the joint probability distributions, panels (b), (d) show the independent joint distributions.

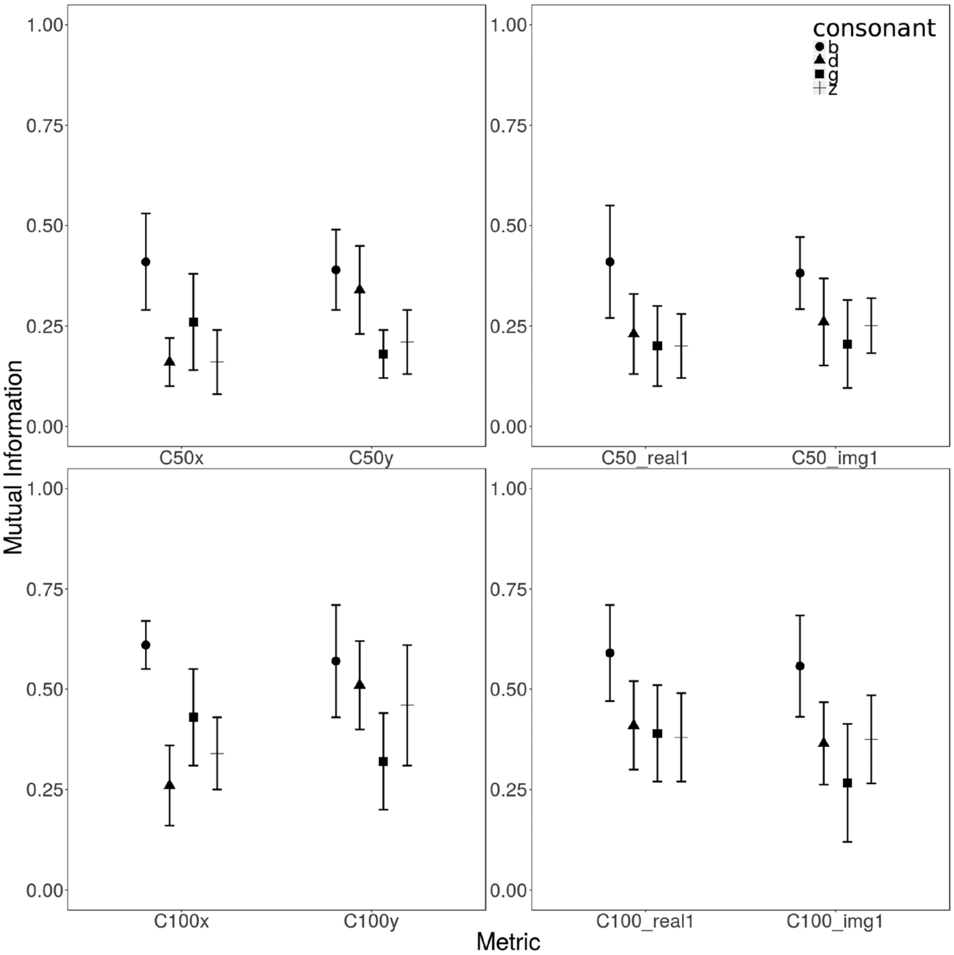

FIG. 3. Mean MI values with bars representing standard deviation for the four consonants /b, d, g, z/ for twelve adult German speakers between the position of the tongue at C50 and V50 (top panels) as well as between V00 and V50 (bottom panels). The tongue position is represented by the following metrics: the horizontal (x) and vertical (y) position of the highest tongue body point (left panels), and the real (1_real) and imaginary (1_imag) components of the first DFT coefficient (right panels).

vowel midpoint (V50) and consonant midpoint (C50) and between the vowel midpoint (V50) and vowel onset (V00), for the vertical and horizontal coordinates of the highest point of the tongue body and for the first DFT coefficient across vowel contexts. Note that the MI values resulting from the measurement of the highest point on the tongue body are divided into the horizontal and vertical components while MI values based on the first DFT coefficient are divided into real and imaginary components.

Table I presents all pairwise comparisons of MI values for the highest point on the tongue body based on the linear mixed-effects model. For the horizontal component of the highest point of the tongue body at C50, the labial stop /b/ shows the highest MI value (0.34) followed by the velar /g/ (0.24), while the alveolar stop /d/ and fricative /z/ have the same value (0.17). There is a significant effect of place of articulation, with /b/ significantly higher than all other consonants ($p < 0.001$), and /g/ significantly higher than /d/ and /z/ ($p < 0.001$), but there is no difference between /z/ and /d/ that both share the alveolar place of articulation.

For the vertical component, the highest value was again observed for the labial stop (0.36) followed by the alveolar stop (0.31) and alveolar fricative (0.21). The velar stop shows the most stability (0.17). Here, the consonants are divided into two groups: /b/, /d/ and /z/, /g/. There is no difference within groups but both /b/ and /d/ are significantly more variable than /g/ ($p < 0.001$ and $p < 1e{-}04$) and /z/ ($p < 0.001$, $p < 1e{-}04$, respectively).

Table II presents pairwise comparisons of MI values for all consonants based on the first DFT coefficient. For the real component of the first DFT coefficient, the labial stop again shows the greatest variability followed by velar stops.

TABLE I. Results of the pairwise comparisons of the MI values for the horizontal (MIx) and the vertical (MIy) position of the highest point on the tongue body between consonant midpoint and vowel midpoint (C50V50)and between vowel onset and vowel midpoint (V00V50) for the four consonants based on the linear mixed-effects model. *P*-values adjusted following Westfall method.

		MIx				MIy			
		β	SE	z	Pr(>|z|)	β	SE	z	Pr(>|z|)
C50V50	d-b	−0.28	0.04	−7.784	<0.001[a]	−0.06	0.04	−1.660	0.184
	g-b	−0.16	0.04	−4.429	<0.001[a]	−0.25	0.04	−6.833	<1e−04[a]
	z-b	−0.28	0.04	−7.873	<0.001[a]	−0.21	0.04	−5.713	<1e−04[a]
	g-d	0.12	0.04	3.355	0.00162[b]	−0.19	0.04	−5.173	<1e−04[a]
	z-d	−0.003	0.04	−0.088	0.92950	−0.15	0.04	-4.053	<1e−04[a]
	z-g	−0.12	0.04	−3.443	0.00162[b]	0.04	0.04	1.120	0.263
V00V50	d-b	−0.36	0.04	−10.247	<0.001[a]	−0.06	0.04	−1.474	0.14048
	g-b	−0.18	0.04	−4.998	<0.001[a]	−0.26	0.04	−6.715	<0.001[a]
	z-b	−0.28	0.04	−7.895	<0.001[a]	−0.12	0.04	−2.979	0.00812[b]
	g-d	0.19	0.04	5.249	<0.001[a]	−0.20	0.04	−5.241	< 0.001[a]
	z-d	0.08	0.04	2.352	0.01867[c]	−0.06	0.04	−1.505	0.13229
	z-g	−0.10	0.04	−2.897	0.00377[b]	0.14	0.04	3.735	<0.001[a]

[a]0 significance.
[b]0.001 significance.
[c]0.01 significance.

TABLE II. Results of the pairwise between consonant comparisons of the MI values obtained for the real (MI_r1) and imaginary (MI_i1) components of the first coefficient of the DFT of the whole tongue contour at consonant midpoint with respect to vowel midpoint (C50V50) as well as at vowel onset with respect to vowel midpoint (V00V50) based on the linear mixed-effects models. *P*-values adjusted following Westfall method.

		MI_r1				MI_i1			
		β	SE	z	Pr(>|z|)	β	SE	z	Pr(>|z|)
C50V50	d-b	−0.20	0.04	−5.431	<1e−04[a]	−0.06	0.04	−1.596	0.350
	g-b	−0.25	0.04	−6.767	<1e−04[a]	−0.06	0.04	−1.650	0.350
	z-b	−0.23	0.04	−6.465	<1e−04[a]	−0.02	0.04	−0.544	0.829
	g-d	−0.05	0.04	−1.336	0.375	−0.001	0.04	−0.054	0.957
	z-d	−0.04	0.04	−1.034	0.375	0.04	0.04	1.052	0.511
	z-g	0.01	0.04	0.302	0.762	0.04	0.04	1.106	0.511
V00V50	d-b	−0.19	0.04	−4.988	<1e−04[a]	−0.20	0.05	−4.326	<0.001[a]
	g-b	−0.20	0.04	−5.442	<1e−04[a]	−0.31	0.05	−6.759	<0.001[a]
	z-b	−0.21	0.04	−5.709	<1e−04[a]	−0.19	0.05	−4.105	<0.001[a]
	g-d	−0.02	0.04	−0.454	0.751	−0.11	0.05	−2.433	0.0217[b]
	z-d	−0.03	0.04	−0.721	0.751	0.01	0.05	0.221	0.8249
	z-g	−0.01	0.04	−0.267	0.790	0.12	0.05	2.654	0.0217[b]

[a]0 significance.
[b]0.05 significance.

However, in this case, the alveolar fricative exhibits a higher MI value than the alveolar stop. Results from LMMs performed on the real component of the first DFT coefficient show that all lingual consonants significantly differ from the labial /b/ ($p < 0.001$), but not from each other. For the imaginary part of the first DFT coefficient the bilabial /b/ shows the highest value followed by the alveolar stop, velar stop, and alveolar fricative, but none of these differences is significant.

At vowel onset (V00), the real component of the first DFT coefficient shows most variability for /b/, followed by /d g z/ that have the same MI value, however, the effect is only significant for the labial stop compared to linguals. As for the imaginary part of the first DFT coefficient, the labial is followed by the alveolar stop and fricative, and the velar is the most stable across vowel contexts. Here, all consonants differ from each other except for the two alveolars. In general, MI values are higher at vowel onset than at consonant midpoint.

Table III presents the results of multiple comparisons based on linear mixed models comparing the MI values between the consonant midpoint and vowel onset for each consonant. Comparisons were made both for the MI values obtained for the highest point on the tongue body and for the first coefficient of the DFT of the whole tongue contours. For both the horizontal and the vertical component of the highest point measure, the MI values are significantly higher at vowel onset than at consonant midpoint. This is also the case for the real component of the DFT metric. For the imaginary component of the first DFT coefficient, all consonants except for /g/ show a significant increase in information sharing at vowel onset as compared to consonant midpoint.

Table IV presents the comparisons based on the output of linear mixed models comparing consonants in terms of the difference in MI values between the consonant midpoint and the vowel onset. None of the comparisons is significant.

IV. DISCUSSION

The main objective of this study was to quantify consonantal differences in CR in German using ultrasound-imaging. To measure different aspects of CR, we employed the measure of MI previously used by Iskarous *et al.* (2013) and compared the variability due to vowel context in the horizontal and vertical position of the highest point of the tongue body with the variability in the position of the whole tongue at two temporal points during consonant production; consonant midpoint and vowel onset. Coarticulatory differences are discussed below with respect to the metric employed: highest point on the tongue body, then, whole tongue contour.

A. The highest tongue body point metric

1. CR differences across places of articulation

To estimate the effect of place of articulation on CR, we measured the highest point of the tongue body. However, the

TABLE III. The pairwise within consonant comparisons of MI values based on the horizontal (Cx) and vertical (Cy) position of the highest point on the tongue body and on the real (r1) and imaginary (MI_i1) components of the first coefficient of the DFT of the whole tongue contour at two time points within the CV syllables: consonant midpoint (Cmid) and vowel onset (Von) for the four target consonant: /b/, /d/, /g/, and /z/. *P*-values adjusted following Westfall method.

		The highest point on the tongue body					The first DFT coefficient			
		β	SE	z	Pr(>|z|)		β	SE	z	Pr(>|z|)
b: Cmid-Von	Cx	−0.48	0.06	−8.395	<0.0000000001[a]	r1	−0.51	0.06	−8.509	<0.0000000001[a]
d: Cmid-Von		−0.12	0.03	3.448	0.000564[a]		−0.20	0.03	5.600	0.000000021384[a]
g: Cmid-Von		−0.19	0.03	5.380	0.00000014868[a]		−0.23	0.03	6.511	0.000000000224[a]
z: Cmid-Von		−0.21	0.03	5.950	0.00000000803[a]		−0.21	0.03	5.912	0.000000006774[a]
b: Cmid-Von	Cy	−0.50	0.06	−7.998	<0.0000000001[a]	i1	−0.48	0.07	−7.281	<0.0000000001[a]
d: Cmid-Von		−0.18	0.04	4.895	0.00000197[a]		−0.12	0.04	2.971	0.00594[b]
g: Cmid-Von		−0.16	0.04	4.422	0.00000976[a]		−0.07	0.04	1.867	0.06186[c]
z: Cmid-Von		−0.27	0.04	7.243	<0.0000000001[a]		−0.14	0.04	3.387	0.00212[b]

[a]0 significance.
[b]0.001 significance.
[c]0.1 significance.

TABLE IV. The pairwise between consonant comparisons of the differences in MI values based on the horizontal (Cx) and vertical (Cy) position of the highest point on the tongue body and on the real (r1) and imaginary (MI_i1) components of the first coefficient of the DFT of the whole tongue contour at two time points within the CV syllable [consonant midpoint (Cmid) and vowel onset (Von)] for the four target consonant: /b/, /d/, /g/, and /z/. *P*-values adjusted following Westfall method.

		The highest point on the tongue body					The first DFT coefficient			
		β	SE	z	Pr(>|z|)		*β*	SE	z	Pr(>|z|)
d(Cmid-Von)-b(Cmid-Von)	Cx	−0.01	0.01	−0.887	0.608	r1	0.04	0.02	1.769	0.288
g(Cmid-Von)-b(Cmid-Von)		0.03	0.01	2.002	0.147		0.01	0.02	0.403	0.914
z(Cmid-Von)-b(Cmid-Von)		−0.00	0.01	−0.229	0.819		0.000002	0.02	0.000	1.000
d(Cmid-Von)-g(Cmid-Von)		−0.04	0.02	−1.769	0.180		0.03	0.02	1.366	0.314
d(Cmid-Von)-z(Cmid-Von)		−0.01	0.02	−0.403	0.687		0.04	0.02	1.769	0.288
g(Cmid-Von)-z(Cmid-Von)		0.03	0.02	1.366	0.314		0.01	0.02	0.403	0.914
d(Cmid-Von)-b(Cmid-Von)	Cy	0.04	0.02	1.769	0.288	i1	0.03	0.03	1.141	0.489
g(Cmid-Von)-b(Cmid-Von)		0.01	0.02	0.403	0.914		0.05	0.03	1.921	0.219
z(Cmid-Von)-b(Cmid-Von)		0.000002	0.02	0.000	1.000		0.02	0.03	0.846	0.637
d(Cmid-Von)-g(Cmid-Von)		0.03	0.02	1.366	0.314		0.02	0.03	−0.780	0.637
d(Cmid-Von)-z(Cmid-Von)		0.04	0.02	1.769	0.288		0.01	0.03	0.295	0.768
g(Cmid-Von)-z(Cmid-Von)		0.01	0.02	0.403	0.914		0.03	0.03	1.075	0.530

highest point metric remains a *local* measure that can only capture the coarticulatory behaviour of the articulator that forms the consonant constriction, i.e., the primary articulator. To address the aspects of CR related to secondary tongue articulators as well as the whole tongue configuration we applied the MI analysis to the examination of the whole tongue contour. Results corroborate previous findings regarding coarticulatory properties of German consonants (e.g., Iskarous *et al.*, 2013; Hoole *et al.*, 1990). For the horizontal position of the highest point of the tongue body, the highest MI values was found for the labial stop /b/ followed by velar stop /g/ and alveolar stop /d/. In the vertical dimension, the labial stop /b/ remained the least constrained of all stops, followed by the alveolar stop /d/, and the velar stop /g/. The latter stop exhibited more resistance to coarticulation with neighboring vowels than all other consonants investigated. This result substantiates previous findings showing that /g/ is more constrained vertically than horizontally (e.g., Fowler and Brancazio, 2000) with a more quantitative dataset. The differences in CR between /b/, /d/, and /g/ are also reflected in the SD of the MI values: while the alveolar stop hardly varies across speakers in the horizontal dimension, variability for the velar stop is minimal in the vertical dimension.

2. Manner-related differences in CR

Regarding the relation between manner of articulation and CR, we expected the alveolar fricative /z/ to exhibit a similar amount of variability to the alveolar stop /d/ in the horizontal dimension but less variance in the vertical dimension. Indeed, while the constriction location for the two alveolar consonants is relatively similar, the constriction degree for the alveolar fricatives production is greater than for the alveolar stop which requires a complete occlusion. Hence, the constriction degree for /z/ must be more precisely controlled than for /d/ to let a certain amount of air pass through the gap between the tongue surface and the hard palate. Results validate our predictions: while no difference between /d/ and /z/ was found in the horizontal dimension, the alveolar fricative was more constrained than the alveolar stop in the vertical dimension.

Overall, results suggest that the highest point metric does capture the CR distinctions between consonants that are related to place and, to a certain extent, manner of articulation.

B. Whole tongue analysis

The results based on the first DFT coefficient provide an opportunity for comparing the local and global aspects of CR. We predicted the DFT-based MI values to generally reflect the highest variability for /b/ in both dimensions, less difference between lingual consonants /d/ and /g/ in the horizontal dimension with the /d/ still being more constrained (based on Iskarous *et al.*, 2013), and the least variability for /g/ in the vertical dimension. A different patterning for /d/ and /z/ was expected due to /z/ being not so constrained in the back of the tongue as /d/.

The MI values based on the first DFT coefficient show a pattern similar to the one provided by the highest point metric; however, here the three lingual consonants pattern relatively close to each other and do not yield any significant differences amongst themselves while all being significantly different from the labial /b/. The fact that the MI value for /d/ is very close to that for /g/ arguably reflects the fact that /d/ is less resistant than /g/ at the tongue back so that this difference offsets the opposite difference in the tongue front. Since the highest point metric captures the constriction location for a given consonant, in the case of /d/ it represents the position of the anterior part of the tongue. Consequently, the information additionally captured by the first DFT coefficient but not present in the highest point metric for /d/ must be provided by the tongue back position.

Regarding the effect of manner of articulation, the MI values based on the real component of the first DFT coefficient for /z/ is lower than that for /d/ while the value based on the imaginary part is higher for /z/ than for /d/. The intersubject variability in the values based on the real component appears to be similar for both consonants whereas the variability in MI values based on imaginary par is higher for /z/. However, none of these differences is significant. This finding suggests that the first DFT component does not reliably capture the larger constraint on tongue dorsum positioning for the fricative production that is associated with the tongue dorsum grooving necessary for the passage of air.

C. Temporal organisation of CR

In light of previous literature addressing the temporal organization of CR within the CV span (e.g., Fowler and Brancazio, 2000), we predicted the information sharing between consonants and vowels to increase from the consonant midpoint to the vowel onset. The prediction was born out: at vowel onset, the MI values increase significantly for all consonants and for all measurements except for the imaginary component of the first DFT coefficient. This increase is more noticeable for less resistant segments such as /b/ and /g/ in the horizontal dimension. However, this difference is not significant. Hence, there was no differences in the temporal evolution of CR between more and less resistant consonants.

There are also changes in the relative patterning of consonants CR between the two time points. The pattern observed for the stops validates our predictions in that the coarticulation degree generally increases over time while the relative positioning of the consonants with respect to each other in terms of their CR is preserved. The fricative /z/, however, patterns differently with respect to other consonants at vowel onset than it does at consonant midpoint. It is not the most resistant consonant anymore, in that it is less resistant than /d/ in the horizontal dimension and less resistant than /g/ in the vertical dimension. This change in CR pattern suggests that /z/ is less constrained temporally than /d/ or /g/.

D. Future directions

The pattern of MI values reported in this study for the highest point of the tongue follows the pattern observed by Iskarous *et al.* (2013) for German stops in the majority of cases. However, a few differences appear. For example, while we find that the voiced labial /b/ is equally constrained in both dimensions, Iskarous *et al.* (2013) observed that its voiceless counterpart is more constrained horizontally than vertically. This may be explained by the fact that voiceless stops are generally more constrained than their voiced correlates due to higher levels of intraoral pressure and less linguopalatal contact (e.g., Recasens, 1999). In the future, it would be useful to compare voiced and voiceless pairs of consonants within one experimental study to advance our understanding of voicing effects on consonants' resistance to coarticulation with neighboring vowels.

While both stimulus sets consisted in series of CVCə nonwords preceded by a schwa and V was stressed, the set of vowels used by Hoole (1999) and analysed by Iskarous *et al.* (2013) included all German vowels, the target sequences were symmetrical and embedded into a longer carrier phrase ("Ich habe geCVCe gesagt"). Finally, target sentences were read as opposed to repeated in our study. This methodological difference may have resulted in some discrepancies in the observed coarticulatory patterns due to differences in orthographic processing in comparison to auditory decoding. Despite these differences, our results show remarkable similarities to previous findings, suggesting that the highest point of the tongue body extracted from ultrasound images captures information similar to that captured by EMA tongue body coils.

In this study, tongue shape was represented by the highest point on the tongue body and the first coefficient of the DFT as *local* and *global* measures of coarticulation. To go a step further and compare the effect of place of articulation requirements on CR with those related to factors affecting the whole tongue, one could employ the MI method with different anatomical subparts of tongue [e.g., in Catalan (Recasens and Rodriguez, 2016, 2017)].

The MI method of quantifying dependence is subject to certain constraints. While it has an important advantage of being purely data-driven rather than affected by specific patterns in the data distribution, this advantage comes at some cost: the MI method requires more data than would be necessary for linear regressions. MI values can also be sensitive to the number of data points depending on the method used for estimation, as is the case with the histogram method used in this study. This issue can be resolved by normalization or by using better methods of estimation, such as kernel density estimation.

Finally, generalization of MI across different experimental techniques (e.g., EMA, ultrasound, EPG, MRI) can contribute to advancing our understanding of speech organization by making the empirical findings of CR research more generalizable across studies and thus directly usable for modeling purposes.

V. CONCLUSIONS

In this study, we tested the reliability of MI in quantifying coarticulatory resistance in German, using ultrasound imaging. Results based on the highest point on the tongue body align with previous findings across languages, that is, an increasing degree of resistance and decrease of MI for syllables involving the labial stop /b/ followed by velar stop /g/ to syllable including the alveolar /d/ and /z/. Hence, MI seems a suitable approach to capture distinctions in consonants′ resistance associated with their place of articulation. Future research should explore other methods of tongue shape quantification as a potential basis for MI calculation to ensure measuring the aspects of CR that involve the whole tongue shape from ultrasound data.

ACKNOWLEDGMENTS

This work was supported by Grant No. DFG GZ: NO 1098/2-1. We are grateful to the LOLA Lab team at the University of Potsdam for their assistance in the collection and processing of the data. We would like to thank Mark Tiede and Jan Ries for developing MATLAB scripts used in this research, and Weirong Chen and Martijn Wieling for their insights regarding the data analysis and statistical analyses. We are also grateful to Carol Fowler as well as to two anonymous reviewers for their helpful comments and suggestions on a previous version of this manuscript. Finally, we would like to thank all participants for making this research possible.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.
Boersma, P., and Weenink, D. (**2015**). "Praat: Doing phonetics by computer," version 6.0.04 [Computer program], http://www.praat.org/ (Last viewed 8/15/18).
Chennoukh, S., Carré, R., and Lindblom, B. (**1997**). "Locus equations in the light of articulatory modeling," J. Acoust. Soc. Am. **102**, 2380–2389.
Darbellay, G. A. (**1999**). "An estimator of the mutual information based on a criterion for conditional independence," Comput. Stat. Data. Anal. **32**(1), 1–17.
Dawson, K., Tiede, M., and Wahlen D. (**2015**). "Methods for quantifying tongue shape and complexity using ultrasound imaging," Clin. Ling. Phon. **30**(3-5), 328–344.
Farnetani, E., and Recasens, D. (**1993**). "Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian," Lang. Speech. **36**, 279–302.
Fowler, C. A., and Brancazio, L. (**2000**). "Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation," Lang. Speech **43**(1), 1–41.
Fowler, C. A., and Saltzman, E. (**1993**). "Coordination and coarticulation in speech production," Lang. Speech **36**(Pt 2–3), 171–195.
Gao, S., Ver Steeg, G., and Galstyan, A. (**2015**). "Efficient estimation of Mutual Information for strongly dependent variables," in *Proceedings of Machine Learning Research: Vol. 38. Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, edited by G. Lebanon and S. V. N. Vishwanathan (PMLR, San Diego, CA) pp. 277–286.
Hoole, P. (**1999**). "On the lingual organization of the German vowel system," J. Acoust. Soc. Am. **106**, 1020–1032.
Hoole, P., Gfroerer, S., and Tillmann, H. G. (**1990**). "Electromagnetic articulography as a tool in the study of lingual coarticulation," FIPKM **28**, 107–122.
Hothorn, T., Bretz, F., and Westfall P. (**2008**). "Simultaneous inference in general parametric models," Biometrical J. **50**(3), 346–363.
Iskarous, K., Fowler, C., and Whalen, D. H. (**2010**). "Locus equations are an acoustic expression of articulator synergy," J. Acoust. Soc. Am. **128**, 2021–2032.
Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E., and Whalen, D. H. (**2013**). "The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance," J. Acoust. Soc. Am. **134**(2), 1271–1282.
Kisler, T., Reichel, U. D., Schiel, F., Draxler, Ch., Jackl, B., and Pörner, N. (**2016**). "BAS speech science web services—An update of current developments," in *Proceedings of the 10th International Conference on Language Resources and Evaluation*, edited by N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis (European Language Resources Association, Paris, France), pp. 3880–3885.
Krull, D. (**1987**). "Second formant locus patterns as a measure of consonant-vowel coarticulation," PERILUS **5**, 43–61.
Liljencrants, J. (**1971**). "Fourier series description of the tongue profile," Speech Trans. Lab. Quart. Progress Status Rep. **12**(4), 9–18.
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. Phon. **40**, 1–19.
Ménard, L., Aubin, J., Thibeault, M., and Richard, G. (**2012**). "Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model," Folia Phoniatr. Logo. **64**(2), 64–72.
Nijland, L., Maassen, B., van der Meulen, S., Gabreels, F., Kraaimaat, F. W., and Schreuder, R. (**2002**). "Coarticulation patterns in children with developmental apraxia of speech," Clin. Ling. Phon. **16**(6), 461–483.
Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., and Tiede, M. (**2018**). "How do children organize their speech in the first years of life? Insight from ultrasound imaging," J. Speech Lang. Hear. Res. **61**, 1355–1368.
Noiray, A., Iskarous, K., and Whalen, D. H. (**2014**). "Variability in English vowels is comparable in articulation and acoustics," Lab. Phonol. **5**(2), 271–288.
Noiray, A., Ménard, L., and Iskarous, K. (**2013**). "The development of motor synergies in children: Ultrasound and acoustic measurements," J. Acoust. Soc. Am. **133**(1), 444–452.
Noiray, A., Ries, J., and Tiede, M. (**2015**). "Sonographic and optical linguo-labial articulation recording system (SOLLAR)," presented at *Ultrafest VII*, Hong Kong.
Öhman, S. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**, 151–168.
Oliphant, T. E. (**2006**). *A Guide to NumPy* (Trelgol, Spanish Fork, UT).
Recasens, D. (**1999**). "Lingual coarticulation," in *Coarticulation. Theory, Data and Techniques*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge), pp. 80–105.
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.
Recasens, D., Pallarès, M. D., and Fontdevila, J. (**1997**). "A model of lingual coarticulation based on articulatory constraints," J. Acoust. Soc. Am. **102**, 544–561.
Recasens, D., and Rodríguez, C. (**2016**). "A study of coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound," J. Phon. **59**, 58–75.
Recasens, D., and Rodríguez, C. (**2017**). "Lingual articulation and coarticulation for Catalan consonants and vowels: An ultrasound study," Phonetica **74**, 125–156.
Roberts, J. K., Monaco, J. P., Stovall, H., and Foster, V. (**2011**). "Explained variance in multilevel models," in *Handbook of Advanced Multilevel Analysis*, edited by J. Hox and J. K. Roberts (Routledge, New York), pp. 219–230.
Scobbie, J. M., Punnoose, R., and Khattab, G. (**2013**). "Articulating five liquids: A single speaker ultrasound study of Malayalam," in *Rhotics: New Data and Perspectives*, pp. 99–124.
Shannon, C. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.
Smith, R. (**2015**). "A mutual information approach to calculating nonlinearity," Stat **4**(1), 291–303.
Sóskuthy, M. (**2017**). "Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction," arXiv:1703.05339.
Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**(3), 1309–1325.
Tabain, M. (**2000**). "Coarticulation in CV syllables: A comparison of Locus Equation and EPG data," J. Phon. **28**(2), 137–159.
Westfall, P. H. (**1997**). "Multiple testing of general contrasts using logical constraints and correlations," J. Am. Stat. Assoc. **92**, 299–306.
Whalen, D. H., Shaw, P. A., Noiray, A., and Antony, R. (**2011**). "Analogs of Tahltan consonant harmony in English CVC syllables," in *Proceedings of the 17th International Congress of Phonetic Sciences*, edited by W.-S. Lee and E. Zee (City University of Hong Kong, Hong Kong), pp. 2129–2132.
Zharkova, N., Gibbon, F. E., and Hardcastle, W. J. (**2015**). "Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation," Clin. Ling. Phon. **29**(4), 249–265.