

The Coarticulation–Invariance Scale:

**Mutual information as a measure of coarticulation resistance, motor synergy,
and articulatory invariance**

Iskarous et al. (2013)

Coarticulation Seminar, WS24-25

Sree Harsha Sunaye

Introduction

Background & Motivation

- **Coarticulation** is the influence of one phonetic segment on another, that causes an aspect of speech to be **variant**.
- However, some aspects of speech remain relatively **invariant**, or un-changing.
- This **Invariance** is seen in aspects that help in *realising* the phonological segment.

Main Motivation:

lack of unified *quantitative approach* used to determine
when an aspect is invariant/variant

Coarticulatory Resistance

- **Definition:** A segment shows resistance to Coarticulatory effects from others.
- e.g.,: Alveo-palatals have *higher* coarticulation resistance than alveolars.

- More Resistance —> Invariance.
- Less Resistance —> Coarticulatory effects.

Now, if only we could measure this Resistance using something, perhaps like a scale...

Coarticulatory Resistance

The Scale

- Scale: Coarticulation and Invariance at opposite ends.
- Measuring Unit? – Mutual Information, MI (more on this later~)

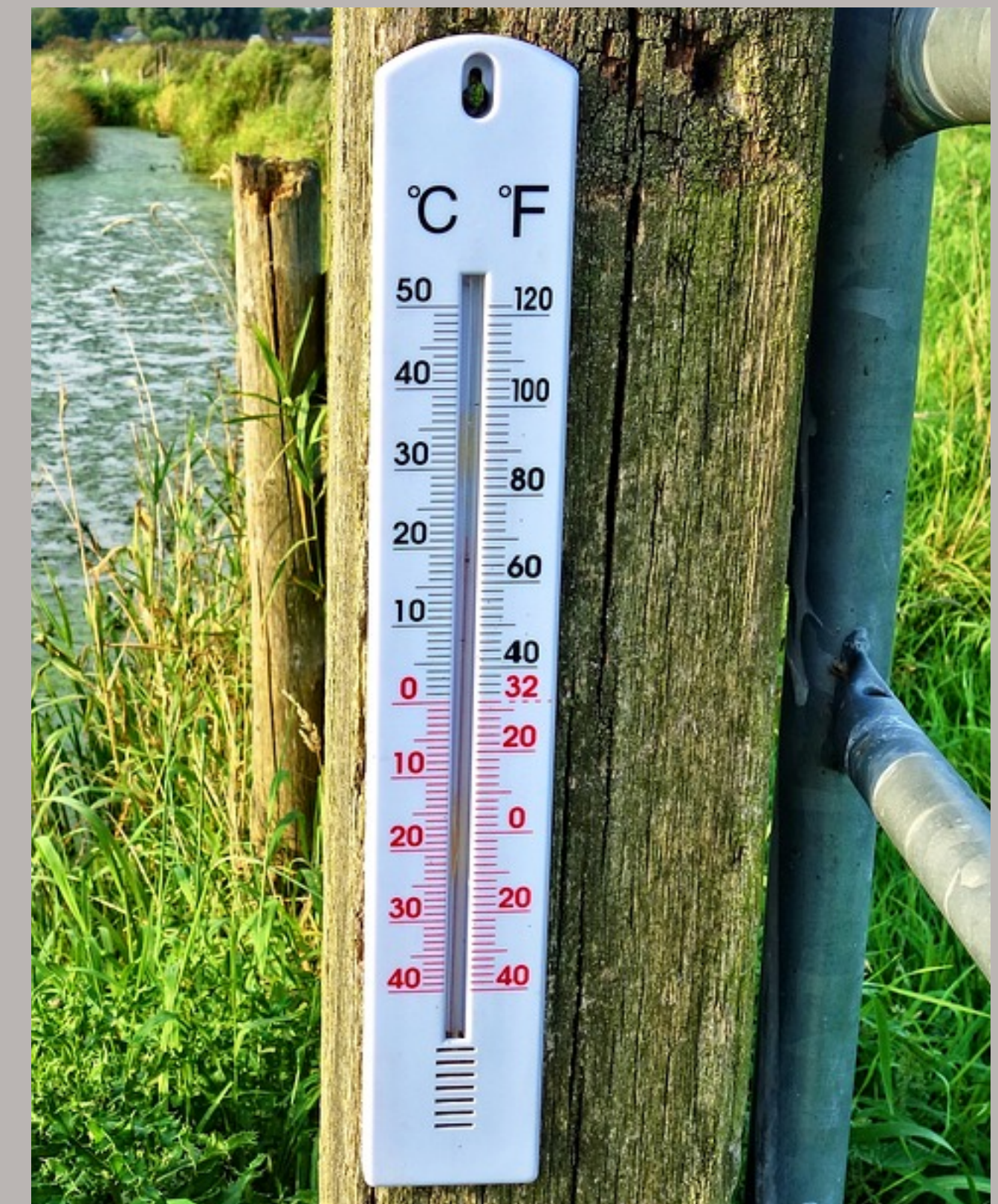
Intuition:

MI values for–

Coarticulation: LARGE

Invariant: small

High end: **maximum** coarticulation



Low end: **minimum** coarticulation

Previous Work

Current work is a development based on 3 Models:

1. Degree of Articulatory Constraint (**DAC**)
2. Locus Equations Model (**LE**)
3. Jackson-Singampalli Statistical ID model.

DAC – Degree of Articulatory Constraint

Recasens and Espinosa (2009)

Aim: Quantify Coarticulatory Resistance and Aggressiveness.

Approach:

- Measure influence of a certain segment on the other.
- Assign a number to this amount of influence; DAC Index
- Higher DAC Index \propto Higher Resistance.

Limitations:

- It is Qualitative – descriptive only.
- Biased towards the tongue and jaw movements.

LE – Locus Equations Model

Sussman et al. (1991); Lindblom and Sussman (2012)

Aim: Quantify Coarticulatory Resistance (Acoustic Domain)

Predictability: of a Consonant's aspects from Vowels.

More Predictable \propto More Coarticulation

Approach:

- Take a Consonant, vary the vowel across all possibilities.
- Plot F2 of a Consonant, against F2 of Vowels.
- Measure? Higher slope = greater degree of coarticulation

Limitations:

- Present only when back of tongue has motion.

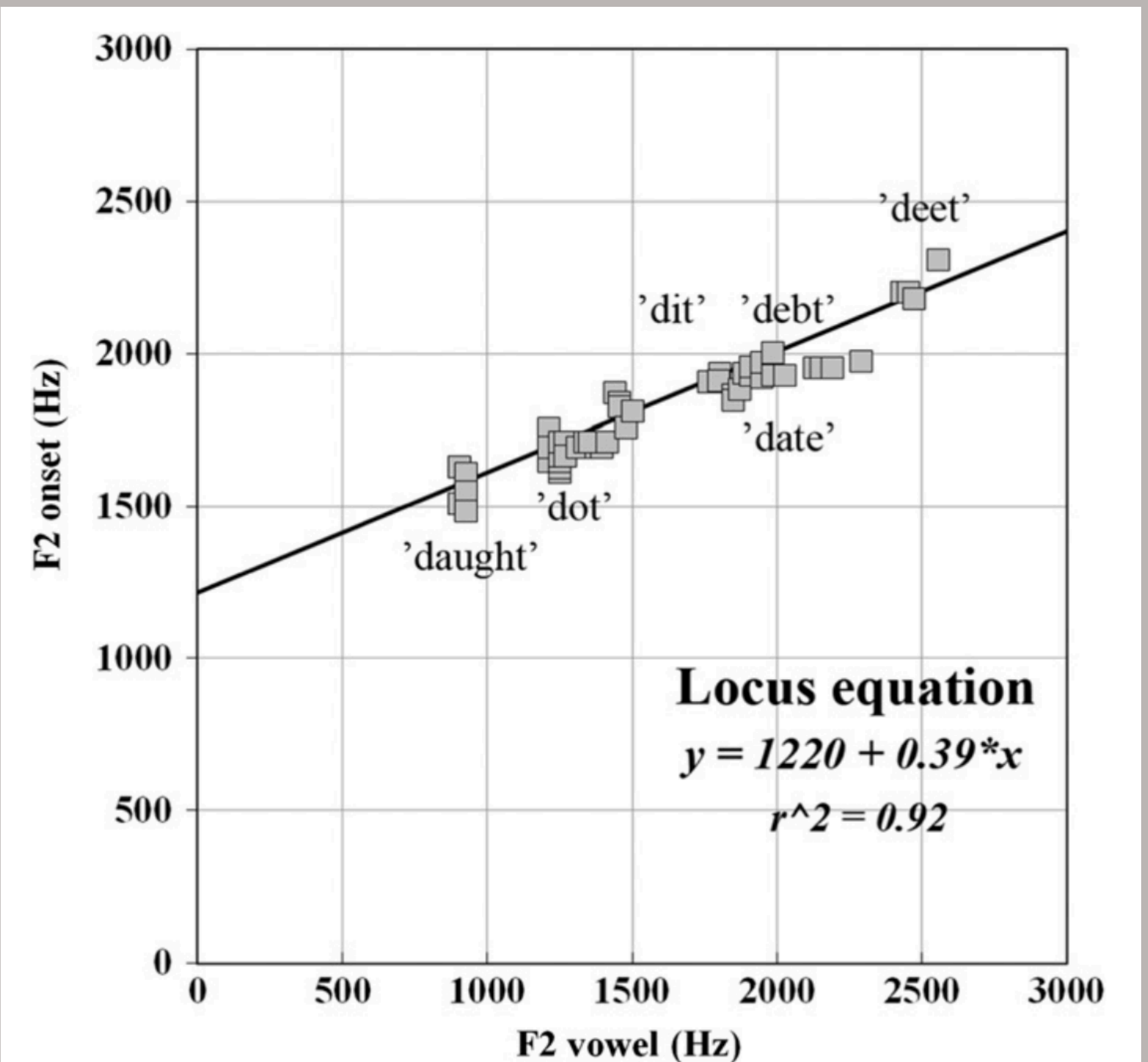


Fig. 1. Representative locus equation scatterplot for [dV] sequences.

Jackson-Singampalli Statistical Identification Model

Recasens (1985), Recasens and Espinosa (2009)

Recasens (1985): Use Standard Deviation (SD) as a measure for Resistance, since it describes the range of a variable around its mean value.

- **Intuition?** SD value tells us how constrained the aspect is.
- More Constrained \propto More Resistant \propto More Invariable.

Problem: No good approach to measure the SD of a variable.

Enter Jackson & Singampalli

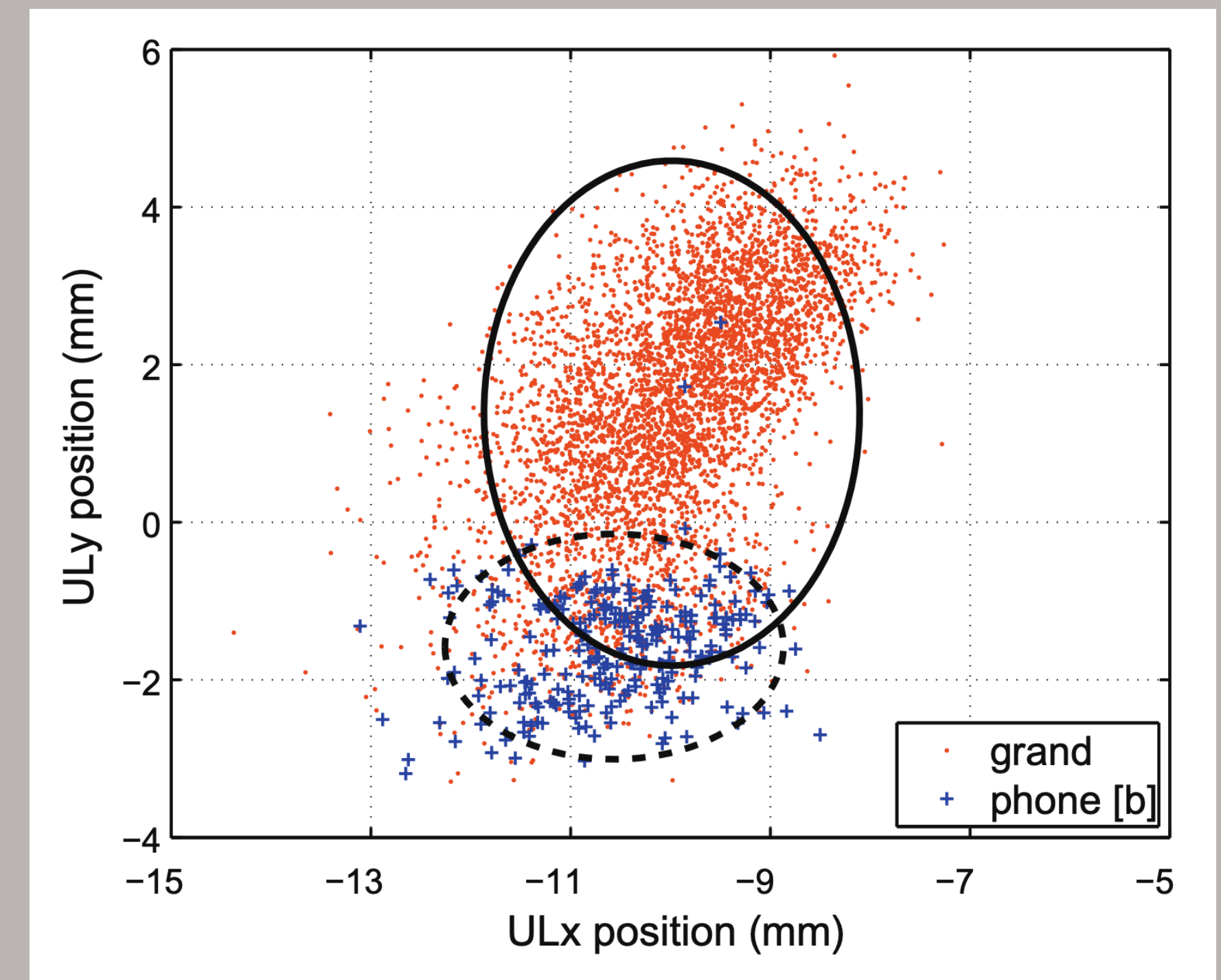
Jackson-Singampalli Statistical Identification Model

Jackson and Singampalli (2009)

Goal: determine how Critical an articulator is for a segment.

For a given Articulator:

- Get Grand & Phone Distributions of a Variable of the articulator, e.g., its Position.
- **Grand:** Everywhere the articulator can possibly be.
- **Phone:** Everywhere the articulator can be, for one segment (phone).
- KL divergence is calculated between Grand & Phone distr. to compare them.



Grand vs. Phone [b] distribution
for Upper Lip

Current Hypothesis

“Mutual Information can be used to define a *Quantitative Scale* with Invariance at the low end and Coarticulation at the high end.”

Mutual Information – MI

Independence and MI

Dependent Variables: when outcomes of one variable, limit the outcomes of the other.

- if two variables are independent: ZERO MI between them.
- if one dependent on another: non-zero MI.

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

MI: Comparing the **Actual Joint Prob.** with **Joint Prob. assuming independence.**

Mutual Information – MI

Procedure

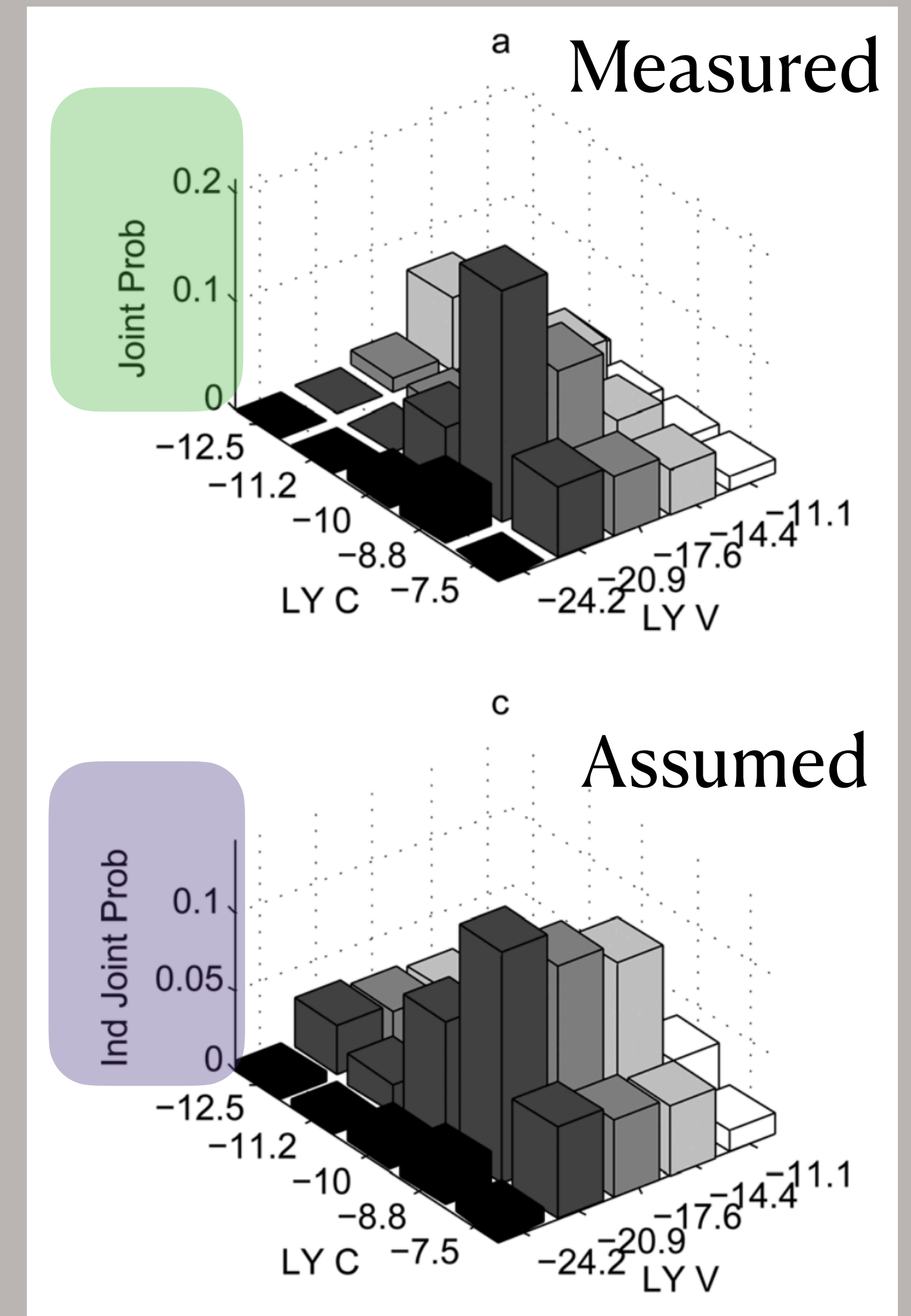
Step 1: Choose a Consonant, and a set of vowels.

Step 2: Choose a Component (e.g., Position) of one of the articulators.

Step 3: Track positions at temporal midpoints of consonant and vowels, in each combination CV.

E.g., Vertical movement (Ly) of lips for /p/ of one speaker.

MI tells us how much the **articulator's position** at the **consonant** can be predicted from that of the vowel.



Datasets Comparison

	German stops	German coronals	Catalan	American English
Aspect in Focus	Place of articulation	Manner of articulation	Place of articulation within Nasals	Temporal Aspect
Segments	Voiceless stops /p,t,k/	Tongue-blade consonants /s, ʃ, t, d, n, l/	vowels /i,a,u/ and /ŋ, n/	/s/ and vowels
Articulators	Tongue, Jaw, Lip	Jaw	Tongue, Lips	Lips, Jaw, Tongue Tip
Equipment	2D electromagnetic midsagittal articulography	2D EMMA	2D EMMA	Taken from XRMB Dataset
Utterance Format	"Ich habe geCVCe gesagt"	"Hab das Verb VCV mit dem Verb VCV verwechselt"	symmetric /pVCVp/	/sVd/
Size	3 stops x 15 vowels	6 consonants x 3 vowels	2 consonants x 3 vowels	1 consonant x multiple vowels

Results

German Stop Consonants: Place of Articulation

Aim: Prove that MI can be used to **distinguish between different places of articulation** by examining MI as a function of segment and articulator.

Graph: MI for Vertical (a) and Horizontal (b) component of stops.

Significantly high MI for effects of place:

Articulator/ Component	L	J	T1	T2&T3	T4
Horizontal	/p,t/	-	/p,k/	/p/	/p/ > /k/
Vertical	/p,t/	-	/p,k/	/p/	/p,t/

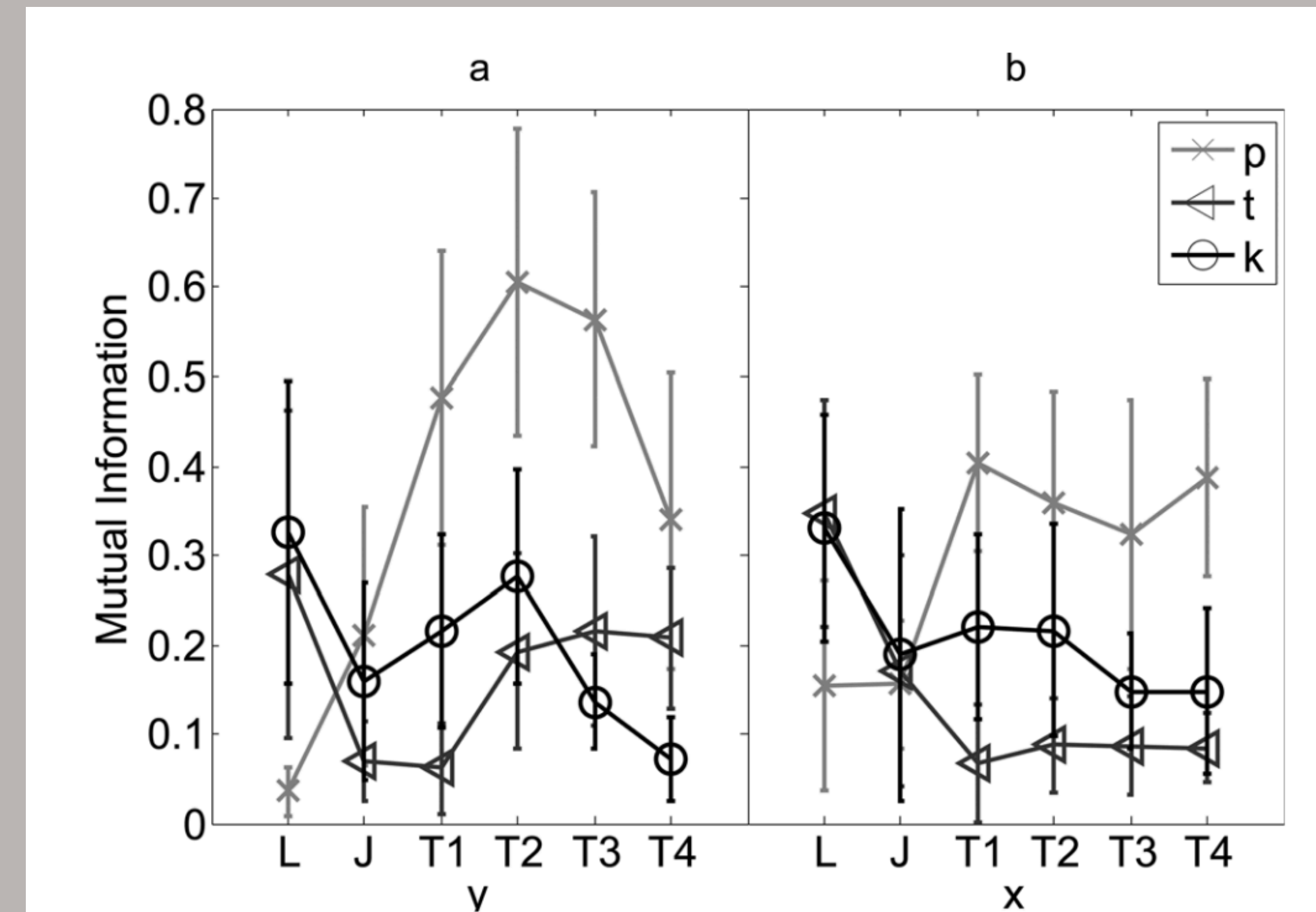


FIG. 2. Mutual information as a function of articulator for three German stops for seven subjects. The error bars are centered at the mean and span a standard deviation on each side of the mean. The data are divided into vertical (a) and horizontal (b) components. L is lower lip; J is jaw; T1 is the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

Results

German Coronals: Manner of Articulation

Aim: Prove that MI can be used to **distinguish between different manner of articulation.**

Graph: MI for Vertical components of coronals

- For tongue, back pellets have higher MI than the jaw and T1
- the sonorants have higher MI than the non-sonorants for the jaw.
- All coronals have low MI for T1.

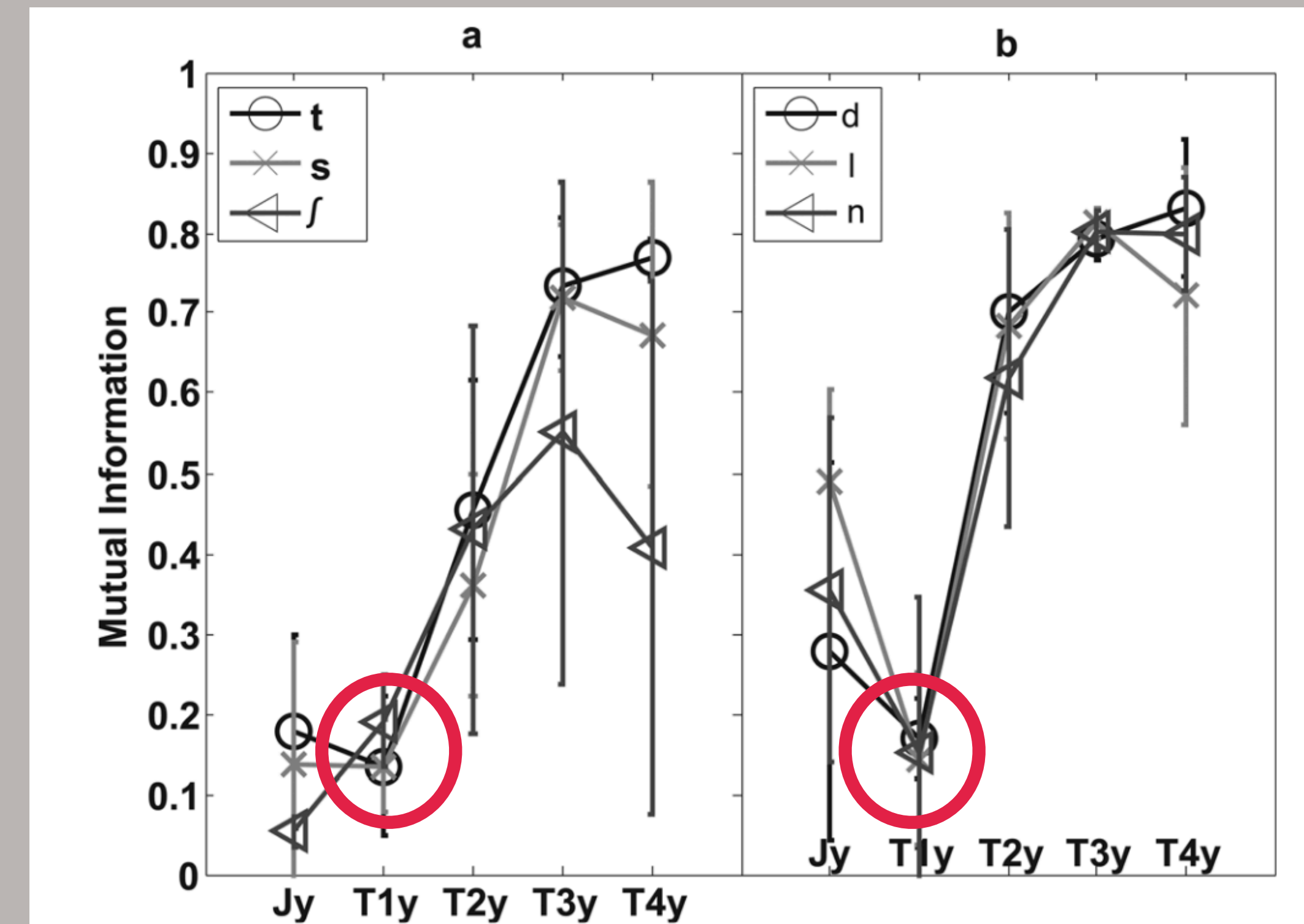


FIG. 3. Mutual information as a function of vertical position of articulators for six German coronals for five subjects. The error bars are centered at the mean for each segment and span a standard deviation on each side of the mean. The data are divided into voiceless coronals /t/, /s/, /ʃ/ (a) and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

Results

Catalan Data: Nasals

Aim: Determine if the difference between palatals and coronals is quantified by MI.

Observations:

- Horizontal: All three tongue sensors showed contrast between two nasals.
- For vertical, only tongue dorsum exhibits this contrast.

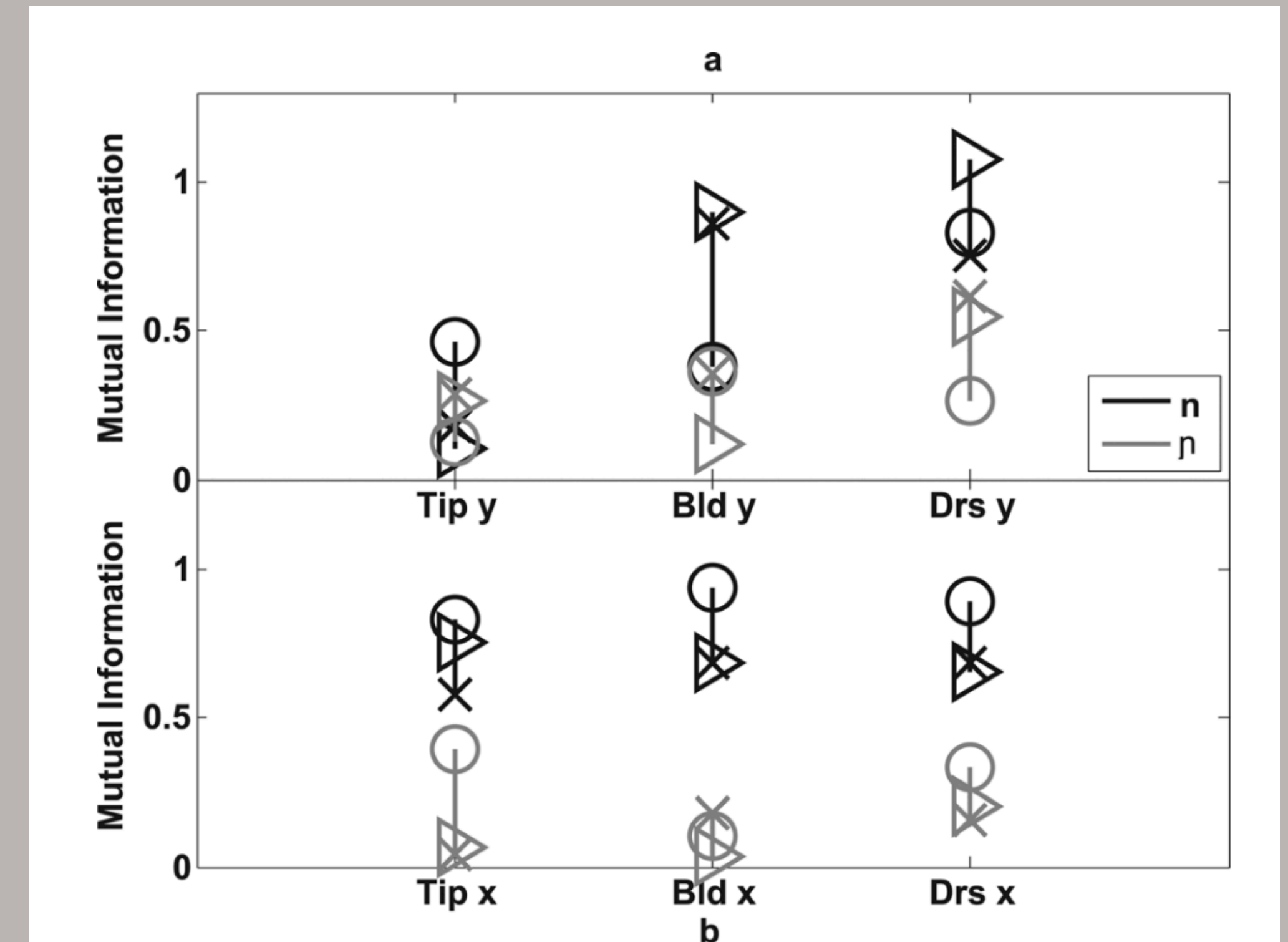


FIG. 5. Mutual information based on V1 as a function of articulator for two Catalan consonants alveolar /n/ and /ɲ/ for three subjects. Labels are as in the original report: tongue sensors for tip, blade and dorsum. Data are divided into vertical (a) and horizontal (b) components. The error bars are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

Results

English data: Time-course of coarticulation

Aim: Can temporal aspect of Coarticulation be described by MI? Effect of time.

Observations:

- horizontal component higher MI than the vertical.
- horizontal: for $t > 2$, all have significant difference of MI to the beginning.
- Vertical: Frames 9 & 10 have significant difference to Frame 1.

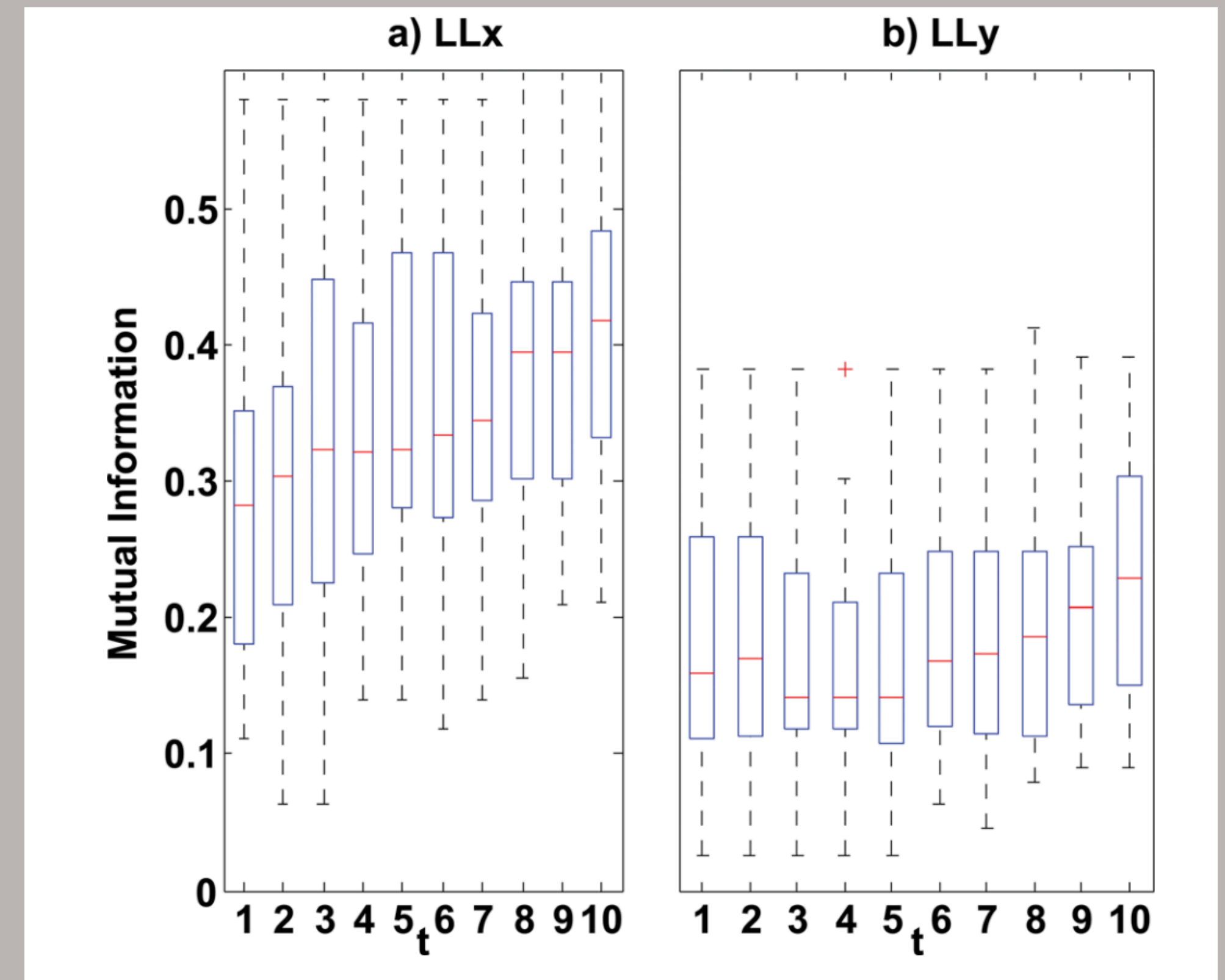


FIG. 6. (Color online) Mutual information as a function of time for the vertical (a) and horizontal (b) components of the lower lip pellet for XRMB data from American English. Variability is across subjects.

Results

English data: Time-course of coarticulation

Graph: Jaw, Tongue Tip

Observations:

- For both articulators, the only significant effects of time on MI are for the last two time frames.

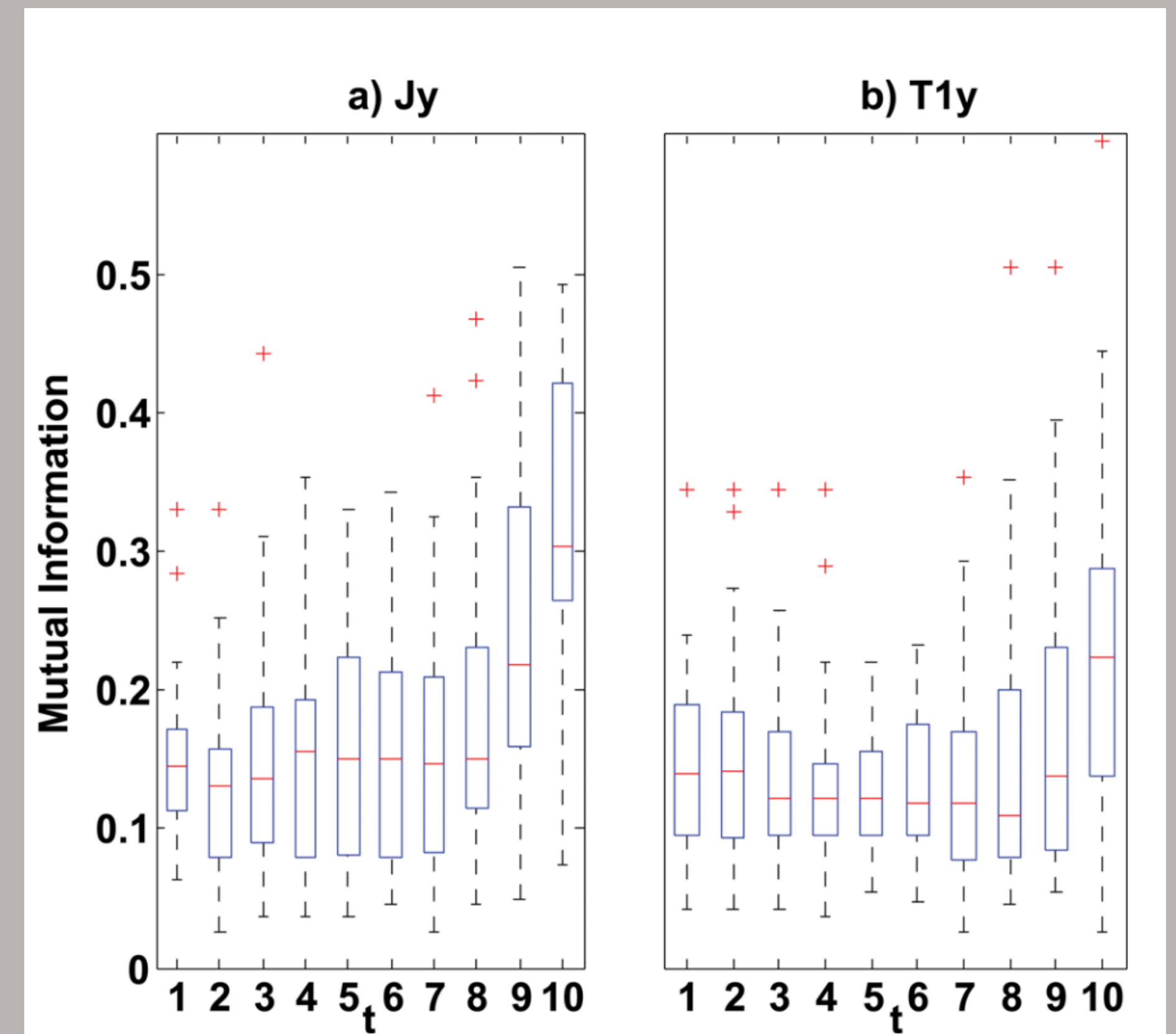


FIG. 7. (Color online) Mutual information as a function of time for the vertical component of the jaw (a) and tongue tip (b) pellets for XRMB data from American English. Variability is across subjects.

Locus Equations and Linearity

In /pV/ sequences:

- the tongue dorsal vertical position for /p/ is roughly linear to that of the vowel.
- Variability of lower lip position is not explained by the vowel.

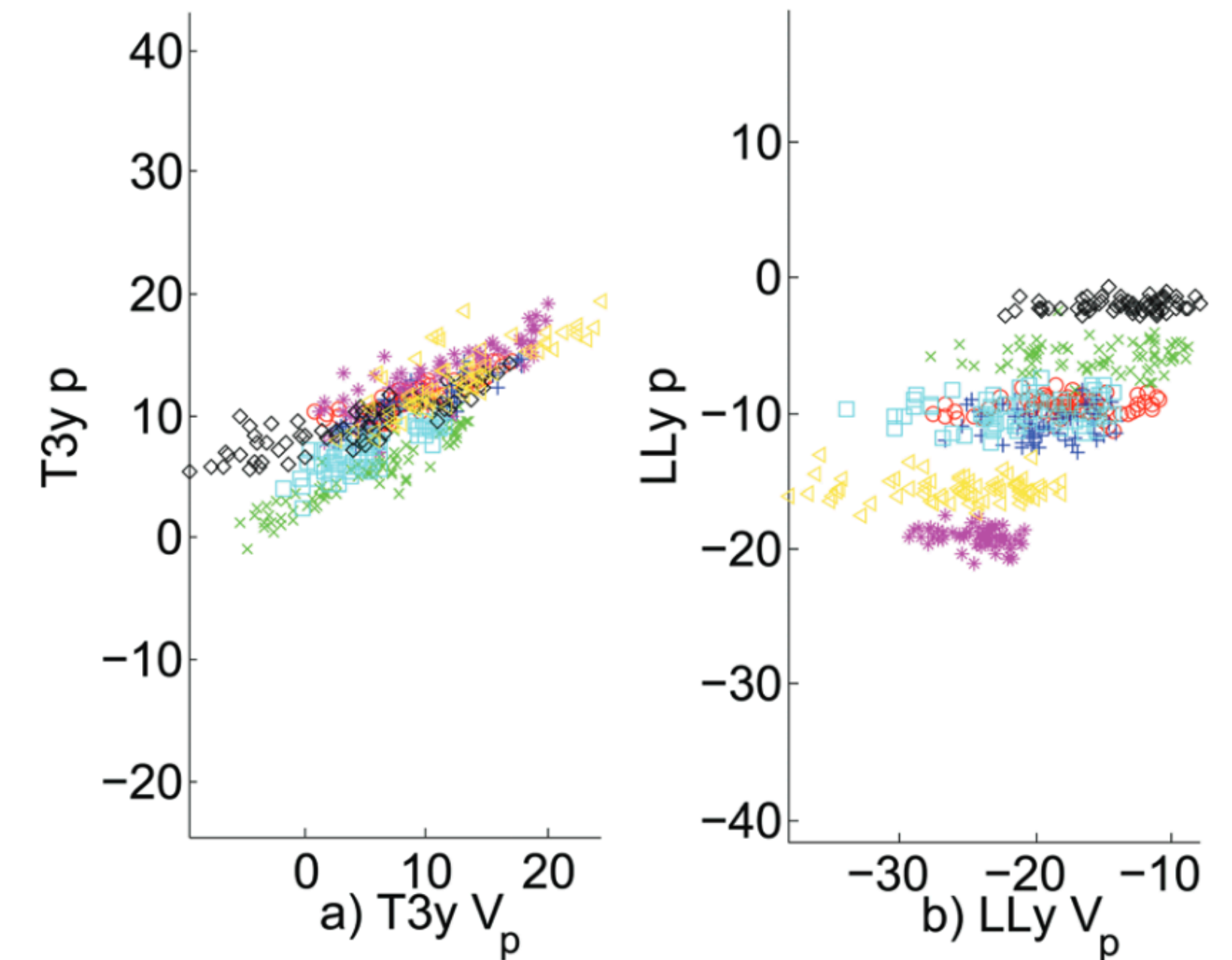


FIG. 8. Scatterplots of the vertical position of the tongue dorsum in the consonant as a function of that for the vowel (a) and for the vertical position of the lower lip (b) in the consonant as a function of the position in the vowel for the German data analyzed in Sec. III. Each color represents a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

Locus Equations and Linearity

MI Graph vs. LE Slope (Dataset 1)

- Overall pattern almost same!
- When MI is large, both the slope and R^2 are high, and the reverse.

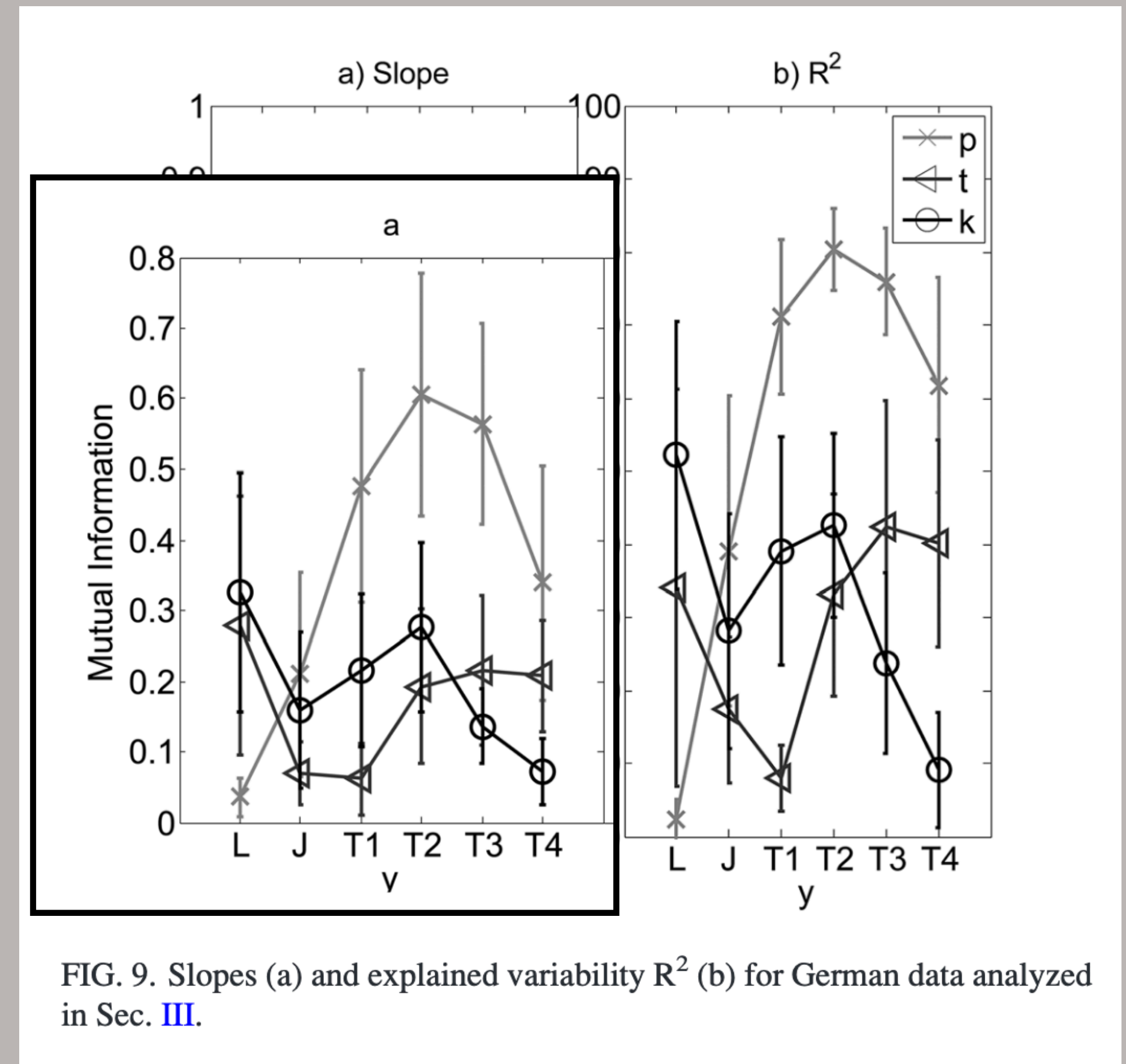


FIG. 9. Slopes (a) and explained variability R^2 (b) for German data analyzed in Sec. III.

Discussion

MI as Coarticulation/Invariance Scale

Dataset 1: German stops

- Labial target is variable for tongue position, which is shown by High MI
- Low MI is observed for components of articulators which are critical for accomplishing that target consonant.
- e.g., for /t/ wrt. vertical component of the tongue and jaw
- For /k/ wrt vertical component of tongue dorsum

Dataset 2: German coronals

- Sibilants and voiceless stop = low MI for Vertical component of Jaw
- This implies Invariance
- All coronals have low MI for tongue tip

Dataset 3: Catalan for Nasals

- Alveo-palatal Nasal has Low MI, implying Invariance

Discussion

MI as Coarticulation/Invariance Scale

Dataset 4: American English for Time-Course Coarticulation

- Proves that MI can be used to understand time-course of Coarticulation.
- e.g., X-axis of Lip shows High MI, much EARLIER than Y-axis.
- /s/ is more invariant during *perturbation* of Aperture, than its protrusion.
- Another example: for jaw and tongue tip, vowel only has an influence on consonant at the END of the sibilant.

Discussion

Limitations of MI

- Comparability of MI across (previous) studies.
- German place of articulation dataset magnitude is half of the manner dataset.
- Generalised description doesn't allow for individual descriptions.

Causes:

- Each study has a different number of subjects/tokens, and different vowels.
- Different placement of pellets on the articulators.

Theoretical Implications

DAC & LE

Property	DAC	MI
Method	Qualitative	Quantitative
Focus of Articulators	Tongue and Jaw	All
No. of Indicators for each Segment	One	Multiple

Locus Equations Model

LE makes ideal assumptions about the distribution, which MI doesn't need to.

Theoretical Implications

Speech Production Theories: Measuring Synergy

Task Dynamics is a production theory; models how articulators move during speech.

- Core idea: Multiple Articulators form a Synergy to produce speech.
- Weight: The amount of Synergy given by each articulator.
- 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.
- Intuition: Weight \propto Resistance (Invariance)

Implications of MI:

- Technical Implication: Speech Synthesis improvement.
- Theoretical Implication: Approaches in alternative research views can be developed together, such as the DAC and LE theories.

Theoretical Implications

Speech Production Theories: Segmental Waves

- **Core Idea:** Each segment has associated “activation wave.”
- This wave represents the degree of *influence* the segment has on vocal tract.
- Time-course of MI is described by Inverse of segmental strength functions.
- MI helps us understand the articulator-independent segment wave better.

So, MI connects this speech theory with LE, DAC, and Motor Synergy.

Conclusions

- Although there are no novel findings compared to the earlier studies, their main contribution is a method to quantify Coarticulation.
- Mutual information can be used as a general measure for characterising the invariance-coarticulation scale.
- This approach captures known generalisations about speech production.
- For both DAC and LE, we do not consider MI analysis to be a falsification of these theories, but to be a generalisation and an extension.
- Relating a set of hitherto unconnected theories of the speech.

Thank you for listening –
Questions?