



**HAL**  
open science

## Speed-curvature relations in speech production challenge the one-third power law.

Pascal Perrier, Susanne Fuchs

► **To cite this version:**

Pascal Perrier, Susanne Fuchs. Speed-curvature relations in speech production challenge the one-third power law.. *Journal of Neurophysiology*, 2008, 100 (9), pp.1171-1183. 10.1152/jn.01116.2007 . hal-00332402

**HAL Id: hal-00332402**

**<https://hal.science/hal-00332402v1>**

Submitted on 23 Oct 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Section:** Neurophysiology  
**Senior Editor:** Karen A. Sigvardt, Ph.D.

**Title:** Speed-curvature relations in speech production challenge the one-third power law

**Running head:** Speech production and the 1/3 power law

**Authors:** Pascal Perrier<sup>1</sup> and Susanne Fuchs<sup>2</sup>

**Affiliations:**

<sup>1</sup>ICP - GIPSA-lab, CNRS, Grenoble Institut National Polytechnique, Grenoble, France

<sup>2</sup>ZAS, Center for General Linguistics, Berlin, Germany

**Corresponding author:**

Pascal Perrier

ICP/GIPSA-lab and INPG

961 rue de la Houille Blanche

Domaine universitaire - BP 46

F - 38402 Saint Martin d'Hères cedex

France

Email: [Pascal.Perrier@gipsa-lab.inpg.fr](mailto:Pascal.Perrier@gipsa-lab.inpg.fr)

Tel: +33 476574825

Fax: +33 476574710

**Number of figures:** 7

**Number of tables:** 3

**Number of pages:** 39

**Number of words for the Abstract:** 249

**Number of words for the Introduction:** 417

**Number of words for the Discussion:** 1128

ABSTRACT

Relations between tangential velocity and trajectory curvature are analyzed for tongue movements during speech production in the framework of the 1/3 power law, discovered by Viviani and colleagues for arm movements. In 2004, Tasko and Westbury found for American English that the power function provides a good account of speech kinematics, but with an exponent that varies across articulators. The present work aims at broadening Tasko and Westbury's study (1) by analyzing speed-curvature relations for various languages (French, German, Mandarin) and for a biomechanical tongue model simulating speech gestures at various speaking rates, and (2) by providing for each speaker or each simulated speaking rate a comparison of results found for the complete set of movements with those found for each movement separately. It is found that the 1/3 power law offers a fair description of the global speed-curvature relations for all speakers and all languages, when articulatory speech data are considered in their whole. This is observed also in the simulations, while the motor control model does not specify any kinematic property of the articulatory paths. However, the refined analysis for individual movements reveals numerous exceptions to this law: the velocity always decreases when curvature increases, but the variation of the slope in the log-log representation is variable. It is concluded that the speed-curvature relation is not controlled in speech movements, and that it only accounts for general properties of the articulatory movements, which could arise from vocal tract dynamics or/and from stochastic characteristics of the measured signals.

Keywords: Motor control, Speech gestures, Speech kinematics

## INTRODUCTION

Studying the kinematic characteristics of speech movements as well as the detailed properties of the articulatory trajectories is an approach that is widely used in speech communication research to infer the underlying control mechanisms of speech gestures. From this perspective, the so-called “*1/3 power law*”, originally proposed by Viviani and Terzuolo in 1982 to characterize the relations between speed and curvature of human movements, is of particular interest. Indeed, since the original study in 1982, numerous investigations have been concerned with the validation of this law and with its potential explanations. These studies greatly contributed to debates about fundamental issues in human motor control which are also crucial for speech communication research, such as the use of kinematic properties in perception mechanisms, or the role of centrally planned optimal motor strategies *versus* physically based factors in the patterning of movement kinematics.

In spite of this potential contribution to speech motor control issues, we are only aware of one study (Tasko and Westbury, 2004) testing the validity of the 1/3 power law for speech movements. Its general observation is that speech movements tend to confirm the 1/3 power law, though with some discrepancies. The authors interpreted their results in terms of articulators’ specific behavior and proposed hypotheses about the underlying motor control. In line with Tasko and Westbury’s work, the present study tests the validity of the 1/3 power law for speech. We will, however, broaden the scope to an analysis of different languages and supplement experimental results with simulations carried out using a realistic model of speech production for which movements are controlled on a target-to-target basis without any specific control of the trajectory shapes or of the velocity profiles. This approach was chosen in order to test the universality of the speed-curvature relations across languages, and to look for possible physical explanations of the 1/3 power law.

Our work is structured in the following way: in the first section, a summary of the main findings related to the 1/3 power law for human movements in general and for speech

production in particular is given. Subsequently, the methodology of our study is presented, involving the description of an experimental study and a modeling approach. In the results-section, the adequacy of the 1/3 power law is tested both for the experimental and the simulated data, and a comparison of these two sets of data is proposed. In the final section, interpretations and conclusions for the 1/3 power law and its potential application to speech motor control are provided.

#### RATIONALE: THE 1/3 POWER LAW

A number of empirical studies have shown that there is a strong coupling between the speed and the trajectory curvature of human movements: when curvature increases, speed decreases and vice-versa. A major formalization of this coupling was provided by Viviani and colleagues (Viviani and Terzuolo, 1982; Lacquaniti, Terzuolo, and Viviani, 1983), who found for planar drawing hand movements that the angular velocity  $a(t)$  and the trajectory curvature  $c(t)$  of the end-effector obey the so called “2/3 power law”

$$a(t) = kc(t)^{2/3} \quad (1).$$

When considering the tangential velocity  $v(t)$  and the trajectory curvature this relation becomes

$$v(t) = kc(t)^{-1/3} \quad (\text{since } v(t) = a(t)/c(t)) \quad (2)$$

which is known as the “1/3 power law”. Factor  $k$  is defined as the “*velocity gain factor*” accounting for differences in average movement velocity. The terms “1/3 power law” and “power law” will be used interchangeably henceforth.

#### *Major findings and potential explanations*

Further evidence supporting Viviani and colleagues’ findings was found for complex arm movements at various rates (Viviani and Cenzato, 1985), for locomotion (Vieilledent et

al., 2001) and for eye motion (de'Sperati and Viviani, 1997), suggesting that the power law could be a fundamental universal characteristic of human movements. It was even suggested that this characteristic was so typical for human movements that it could strongly influence human perception in terms of naturalness and classification (Viviani and Stucchi, 1992) as well as in terms of perceptual anticipation of impending events (Kandel et al., 2000).

There have been many debates in the literature about the possible origin of the relation between speed and curvature, and whether or not it reflects the underlying motor control of the central nervous system. It has been suggested that the power law results from optimization principles underlying human movement production, such as jerk minimization (Viviani and Flash, 1995), maximum smoothness principles (Todorov and Jordan, 1998) or minimizing the impact of neural noise on target reaching accuracy (Harris and Wolpert, 1998). In contrast, it was also proposed that the 1/3 power law could arise from biomechanical properties of the peripheral motor system (Gribble and Ostry, 1996). More recently, Maoz et al. (Maoz et al, 2005) suggested that it could, at least partly, arise from the noise that is inherently added to the kinematic data, because of measurement inaccuracy and/or the stochastic characteristics of the motor system (including neural noise).

Arguments against the hypothesis of a major role for physical factors in the power law are mainly based on the work of Massey et al. (Massey et al., 1992), who demonstrated that the same coupling between speed and curvature still held true when subjects were drawing curved patterns in isometric conditions with a joystick, i.e. when the motor system did not move. However, Gribble and Ostry (Gribble and Ostry, 1996) moderated the relevance of this finding, since they showed by means of their model that under isometric conditions force variations also obey the same law. Another concern about the viability of the biomechanical explanation comes from the fact that the power law was found in mechanical systems as different as upper limb, lower limb, and eyes. A possible answer to this can be found in Gribble and Ostry (1996) and Schaal and Sternad (Schaal and Sternad, 2002). The former

suggest that basic spring-like characteristics, shared by all motor systems, would justify the impact of biomechanics. The latter assume that these spring-like characteristics naturally ensure the smoothness of the trajectory necessary for the power law to apply.

Criticism of the hypothesis of a dominant influence of centrally planned factors in the power law is mainly based on the observation that many exceptions to this law exist in complex human movements. Schaal and Sternad (Schaal and Sternad, 2002) observed for instance that during the drawing of large 3D elliptical patterns, the exponent of the non-linear relation between speed and curvature often greatly deviates from the theoretical value (-1/3). They concluded that the 1/3 power law should be seen as “*an epiphenomenon of smooth oscillatory trajectory generation in joint space*” (p.16).

#### *Preliminary results for speech production*

This summary suggests that, despite its controversy, the 1/3 power law is a major experimental finding that raises fundamental issues about the control and perception of human movements: What kinematic properties of human movements are directly related to centrally planned factors? To what extent are properties of the motor system used to perceive human movements?

In this context, speech movements should have been of particular interest. Indeed, speech biomechanics is very specific, since two of the main vocal tract articulators (tongue and lips) are non-rigid bodies interacting with hard structures (a complex biomechanical case to study), and since time variable boundary conditions (tongue-palate and tongue-teeth interaction; upper lip-lower lip and lip-jaw interactions) probably constrain speech gestures to a great extent. In addition, speech task specification is likely to vary across the worlds' languages because of the well-known differences in size and complexity of phonological inventories. Moreover, speech gestures in the oral cavity are to a large extent not perceived

visually, but auditorily via the emitted speech signal after a non linear transformation from the articulatory to the acoustic domain.

In spite of these interesting specificities, to our knowledge, only one study has addressed the power law issue in the context of speech production. It was carried out by Tasko and Westbury in 2004 who took advantage of the large X-ray microbeam speech production database to analyze the kinematic characteristics of speech for 18 American English speakers (Tasko and Westbury, 2004). They found an exponent value which was near, but not exactly (-1/3). The exponent value, the velocity gain factor, and the strength of the speed-curvature relation varied systematically across subjects and pellets. For the whole set of data, the exponent values ranged from -0.44 to -0.34. Therefore, the absolute values of the exponent are slightly above the range expected from the original power law for planar drawing movements. However, they are much smaller than the absolute values found by Pollick and Ishimura (Pollick and Ishimura, 1996) for 3D straight-ahead point-to-point movements (ranging from 0.52 to 0.67). According to Tasko and Westbury, this range of variation in the exponent values suggests that speech gestures are closer to planar drawing movements for which the whole trajectory is controlled than to target directed movements (p.77).

On average, the variance of speech velocity explained by the power law ranged from 65% to 70%, which is smaller than the values observed for drawing and visual tracking movements. This suggests a weaker speed-curvature coupling in speech movements than in other human movements. The authors ascribe this to the fact that their speech material included consonantal segments, where the tongue is in contact with the palate, which constrains its movement. Finally Tasko and Westbury proposed that interarticulator differences found in the exponent and in the velocity gain factor values reflect the behavioral specificities of each orofacial structure.



The present work aims at broadening Tasko and Westbury's study in essentially two directions: (1) the potential influence of language specific constraints is assessed by studying the speed-curvature relationships for speakers of relatively unrelated languages (French, German, Mandarin Chinese); (2) the contribution of biomechanical factors is tested by analyzing the speed-curvature relationship in artificial tongue movements generated with an anthropomorphic speech production model in comparison to the experimental data. In both cases, special attention has been devoted to the analysis of the power law adequacy at the level of each singular tongue gesture. Indeed, if Viviani and colleagues' hypothesis is true, and if the power law results from motor control strategies that are so specific to human skilled movements that they can be perceptually identified, it should be systematically observed for each single gesture. Hence, in our study speech sequences were splitted into small movement segments delimited by two successive velocity zero crossings, and speed-curvature relations were calculated in parallel for the whole set of data and for each segment separately.

## MATERIALS AND METHODS

### *Experimental data*

In order to evaluate the universality of the power law, it was decided to analyze articulatory data from various languages and various speakers, collected in the context of studies with various objectives. These data were gathered at ICP in Grenoble and at ZAS in Berlin in the last ten years. Similar experimental set-ups were used in both labs (2D Electromagnetic Midsagittal Articulograph AG100 from Carstens Medizin Electronics), with the same data post-processing<sup>1</sup>, and with similar experimental protocols.

### **Subjects**

Three rather unrelated languages were selected: French, German, and Mandarin Chinese. Two speakers were considered for each language (F1 and F2 for French, G1 and G2 for German, C1 and C2 for Mandarin Chinese). They were male subjects except F1. Their

ages ranged from 25 to 30 except for C1 who was 45. They were scientists working in our labs except C2. All the speakers had no history of speech, language or hearing pathologies. The French and the Mandarin Chinese data were collected in the context of a cross-linguistic study of coarticulation mechanisms (Ma et al., 2006), while the German data were obtained for a study of interspeaker token-to-token variability (Mooshammer et al., 2004). Details about the experimental procedures and the data processing are given in Ma et al. (Ma et al., 2006) and Mooshammer et al., (Mooshammer et al., 2004). Only the most immediately relevant aspects of the method are described below.

### **Speech Tasks**

The cross-linguistic study of coarticulation of Ma et al. (Ma et al., 2006) aimed at characterizing in Vowel1-Vowel2 and Vowel1-Consonant-Vowel2 sequences how the production of the second vowel Vowel2 influences the articulation of the preceding sounds (*anticipatory coarticulation*). The authors of this study were especially interested in looking at potential interactions between the phonological properties of the language and anticipatory coarticulation strategies. Mandarin Chinese and French were chosen because the phonological status of the syllable differs across the two languages. The same sequences were recorded for both languages. They were carefully chosen in order to respect the phonotactical rules of each language. In Mandarin Chinese all syllables were pronounced with a high level tone (i.e. flat tone). The vowels were one of /i, a, u/ and the consonant one of /t, k/. The sequences were recorded both as nonsense words in meaningful carrier sentences (example for French “C’est **aki** ça ?”, i.e. “Is that aki?”) or as parts of meaningful words or groups of words within meaningful sentences (example for French “C’est Harry qui **t’a touché** ?”, i.e. “Did Harry touch you?”). For the current study, we selected the latter part of the corpus for both languages, because it offers a larger variety of sounds. Ten repetitions of five different sentences lasting for around one second each were considered for subjects F1, F2 and C2. For

subject C1, who was the subject for the pilot study, the corpus was slightly different and corresponds to four repetitions of 20 different sentences.

The German material was collected in order to study the patterns of articulatory variability measured for each speaker in different repetitions of a vowel in the same phonetic context. More precisely, the authors analyzed the potential influences of perceptual constraints, morphological characteristics of the vocal tract and linguistic factors on the patterns of variability. In its whole the corpus consisted of C1VC2/ə/ nonsense words with C1 and C2 being either the velar stops /k/ and /g/, or the bilabial stops /p/ and /b/, and V being one of the 14 tense or lax German vowels. The initial stop C1 was phonologically voiced and the medial C2 was phonologically voiceless. V is in a stressed position. All nonsense words were embedded in the carrier sentence “Sage ...bitte” (“Say .... please”) and they were repeated 10 times. The whole set of data yet consisted of 280 sentences for each subject. For the current study, a limited subset of these data was analyzed, consisting of 50 randomly selected sentences for each speaker.

### **Data Acquisition**

For all speakers, the articulatory data were collected with the AG100 system. With this device measurements are based on the electromagnetic induction phenomenon. Alternating magnetic fields are generated by three transmitter coils that are mounted on the corners of a helmet that has the form of an equilateral triangle. Small transducer coils (called henceforth *sensors*) are attached to the articulators in the midsagittal plane. Each sinusoidal magnetic field induces, because of its time variation, a sinusoidal current in each sensor, the intensity of which is proportional to the inverse of the cube of the Euclidian distance separating the transmitter and the sensor. The measurement of the intensity of the three currents generated in each sensor by the three transmitters, after some corrective pre-processing, yields measures of this sensor’s locations as a function of time. In order to minimize measurement errors, the sensors have to be accurately located in the mid-sagittal

plane of the head. Two sensors, located on fixed parts of the head, are used as reference sensors to compensate for possible head rotation in the mid-sagittal plane. For each measurement, a tilt factor is calculated that tells whether or not the measurements can be trusted. Further information about this technique can be found in Perkell et al. (Perkell et al., 1992) and in Hoole and Nguyen (Hoole and Nguyen, 1999).

----- Insert Figure 1 around here -----

For all subjects, 4 sensors were located on the tongue (note that for speaker F2 the Tback sensor stopped working during the experiment; it was not all considered for this study) . They were quasi evenly distributed from approximately 1 cm to 5 cm from the tongue tip. Figure 1 (top panel) shows a schematic view of these sensors locations on the tongue contour in the mid-sagittal plane. Sensors were designated (from front to the back) as Ttip, Tblade, Tdors, and Tback. Reference coils were glued to the bridge of the nose and above the upper incisors.

The horizontal direction of the speaker's bite plane was also measured at the end of each session. To do so a plexiglass sheet was inserted in the subject's mouth. Two sensors were located on this sheet along the anterior-posterior direction. The line joining these sensors, when the speaker is asked to bite against the sheet, is considered to be the horizontal direction. The articulatory data were then rotated around the reference sensor on the upper incisor, in order to ensure that the x and y directions of the rotated data correspond respectively to the speaker specific front/back and high/low directions. Indeed, these directions are essential for the phonetic characterization of speech movements (see for example Ladefoged and Maddieson, 1996)

Depending on the subject, data were sampled either at 200 Hz (subjects C1, C2, G1 and G2) or at 500 Hz (subjects F1 and F2). The articulatory signals were low pass filtered (see section *Data analyses*). Examples of trajectories of the tongue dorsum sensor (3<sup>rd</sup> sensor from

left in Figure 1, upper panel) after filtering are given in Figure 1 (lower panel) for randomly selected time windows of 1 s, for speakers F1, C1 and G2.

### *Simulated data*

To study the potential impact of biomechanical factors on the speed-curvature relation, tongue movements were simulated with a biomechanical model controlled on a target-to-target basis (Payan and Perrier, 1997; Perrier et al., 2003). Earlier works in our group (see in particular Payan and Perrier, 1997; Perrier et al., 2003; Perrier et al., 2005) have demonstrated that the tongue trajectories simulated with this model are realistic and similar to real trajectories measured on human speakers.

In this model, the kinematic properties of the movements are not specified at the motor control level. Tongue movements are generated as sequences of sub-movements between targets. In the complete speech production model, called *GEPPETO*, these targets are related to the phonological structure of the speech sequence (Perrier et al., 2005): targets are considered as physical correlates of the phonological inputs. However, in the current study, this motor control layer was not used, and the successive targets underlying the production of tongue movements are just considered to be intermediate spatial objectives within a larger tongue movement, without any explicit links to the phonological level.

For the production of a movement, the motor control system specifies via the motor control variables (see below) the mechanical properties and the timing (in terms of transition time between targets and hold duration of each target) of the successive targets. The shift of the motor control variables between the successive target values is made at a constant rate and its timing is setup independently of any consideration related to the shape of the articulatory path or to articulatory kinematics. Thus, in this model, the kinematic properties of the movements (in particular the trajectory curvature and the tangential velocity) are neither explicitly controlled nor the consequences of an optimal motor control strategy. The kinematic properties are, on the contrary, the results of a combination of central and

peripheral influences, namely those of the motor command values and their timing, those of the muscle force generation mechanisms (see below), and those of the dynamical and anatomical properties of tongue tissues (inertia, elasticity, muscle fiber arrangements).

----- Insert Figure 2 around here -----

In this model, elastic properties of tongue tissues are accounted for by finite-element (FE) modeling with a mesh defined by 221 nodes and 192 elements. The FE mesh represents a 2D projection of the tongue in the mid-sagittal plane. It is inserted inside a projection of the vocal tract in this plane, characterized by curves representing the contours of the lips, palate, and pharynx. The jaw and the hyoid bone are also represented in this plane by static rigid structures to which the tongue is attached. Mechanical contacts of the tongue with the palate and the teeth are also modeled. The main muscles responsible for shaping and moving the projection of the tongue in the mid-sagittal plane are taken into account: posterior and anterior parts of the genioglossus, styloglossus, hyoglossus, inferior and superior longitudinalis, and verticalis. They are represented in the model at two different levels (see Figure 2). First, their action on the tongue body is accounted for by “macro-fibers” (the bold lines on Figure 2) that specify the direction of the forces and the nodes of the FE mesh to which the forces are applied. Second, since the activation of a muscle modifies the elastic properties of the muscle tissues, muscles are also represented in the model by a number of selected elements within the FE structure (gray shaded elements on Figure 2), whose mechanical stiffness increases with muscle activation.

Muscle force generation mechanisms are modeled in a functional way according to Feldman’s Equilibrium Point Hypothesis of motor control (Feldman, 1986). This model reflects the claim that  $\alpha$  motoneuron activation, which generates force, is not centrally controlled, but is the consequence of the interaction between a central command (called  $\lambda$ ) and afferent inputs related to muscle length and velocity. The activation is zero if the muscle length is shorter than  $\lambda$ . Otherwise, the activation is a function of the difference between the

muscle length and  $\lambda$ , and of the muscle length change rate. The relation between active muscle force and muscle activation is approximated by an exponential function. In addition, the *sliding filament theory* is taken into account in the model: the force generation capability of a muscle depends on muscle shortening or lengthening velocity (see Payan and Perrier, 1997 for more details).

For the purpose of this study, 600 sequences consisting of an initial schwa (corresponding to the tongue position at rest) followed by 3 different articulatory targets were generated. The motor control variables used for the 3 articulatory targets following the initial schwa were selected using a Monte-Carlo method based on a uniform sampling of the motor command space. Consequently, the targets used in the simulations correspond to tongue shapes that are randomly selected among all the possible tongue shapes that can be generated by the model, without any intention to replicate the corpora used for the experimental data and without any attempt to make any link between these targets and the phonemes of a given language. Movements can include parts where the tongue is in contact with the palate, but for the majority of the sequences this did not happen.

In order to also evaluate to which extent the velocity gain factor of the power law accounts properly for movement velocity variations, simulations were carried out for 3 speaking rates: normal, fast, and slow. These sequences will be referred to henceforth as *simulated data* in comparison to the experimental data. The time courses of the motor control variables were defined by a hold duration, for which the motor commands at targets were kept constant, and by a transition duration, for which motor commands were shifted at a constant rate from one target value to the next. At slow speaking rate the hold duration and the transition duration were both 120ms, at normal speaking rate they were respectively 100ms and 80ms, and at fast speaking rate they were both 60ms. Durations at normal speaking rate were chosen because they allowed generating movements with durations and velocity amplitudes consistent with experimental data.

The trajectories of 3 flesh points located on the tongue surface in the model were considered for analysis, similar to the tongue tip, tongue blade, and tongue dorsum sensor in the experimental data. In addition, we also considered a flesh point at the surface of the tongue located in the pharyngeal region, since movements from this region could not be obtained on the basis of the experimental data. The trajectories for the simulated sequences were sampled at 125 Hz.

### *Data analyses*

To analyze the relation between speed and curvature, the same methods were applied to experimental and simulated data. The data were first low-pass filtered with a linear phase Remez filter (Gain 0 dB from 0 to 15 Hz; Gain < - 46dB above 30 Hz). As shown by Maoz et al. (Maoz et al., 2005), such a low-pass filter has the advantage of potentially decreasing the contribution of noise to the emergence of the power law. After filtering, the first and the second derivatives ( $\frac{dx}{dt}$  and  $\frac{d^2x}{dt^2}$ ) were computed for each sample using the finite differences approximation. To be accurate, this method requires a sampling frequency that is extremely large as compared to the bandwidth of the signal (15 Hz for our data). To ensure accuracy, the low-pass filtered data were oversampled by a factor 10 (corresponding to a sampling frequency varying from 1250 Hz for the simulated data to 5000 Hz for the experimental data collected from the French speakers) before the derivatives were calculated. Derivatives were used for calculating tangential velocity  $v(t)$  as follows:

$$v(t) = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \quad (3).$$

The curvature was computed as the absolute value of the Frenet-Serret formula (Viviani and Stucchi, 1992):



$$c(t) = \left| \frac{\left( \frac{d^2 x}{dt^2} \frac{dy}{dt} - \frac{d^2 y}{dt^2} \frac{dx}{dt} \right)}{v(t)^3} \right| \quad (4).$$

The power law model makes use of a velocity gain factor accounting for variability in average velocity across sequences. Average velocity is likely to vary across sentences and across phoneme-to-phoneme transitions within sentences. To take this possibility into account, the data were analyzed in three different kinds of sets:

- a. *All data (All)* for each subject and each sensor in the real sequences and for each sensor and each speaking rate condition in the simulated sequences.
- b. Data were further split by *sentence (Sent)*. A sentence corresponds to an actual carrier phrase in the experimental data, and to a simulated sequence (i.e. the initial schwa followed by 3 different articulatory targets) for the simulated data.
- c. Data were further split by *segment (Seg)*. To determine a segment, a low velocity threshold  $V_{min}$  was specified and it was assumed that movements are only realized when the tangential velocity is higher than  $V_{min}$ . Thus, a segment corresponds to a part of the signal during which the tangential velocity is consistently higher than  $V_{min}$ . Segments can then be considered as continuous movements separated by quasi steady state articulatory positions. The threshold value  $V_{min}$  was set to 0.5 cm/s. In addition, only the segments corresponding to a global displacement (characterized by the length of the whole path described by the sensor) larger than 1.5mm were taken into consideration. This minimum threshold of 1.5mm was chosen because it corresponds to transitions between close sounds such as /i/ and /e/ measured for real speakers in the articulatory space (see for example Ma et al., 2006).

For each dataset, log-log linear regression coefficients were calculated between the decimal logarithms of  $v(t)$ ,  $\log_{10}(v(t))$ , and  $c(t)$ ,  $\log_{10}(c(t))$ , according to the formula:

$$\log_{10}(v(t)) = a \log_{10}(c(t)) + b \quad (5).$$

Coefficient - a - corresponds to the exponent of the power law and coefficient - b - to  $\log_{10}(k)$ , where - k - is the velocity gain factor. The log-log linear regression approach was criticized by Schaal and Sternad (Schaal and Sternad, 2001), who showed for high velocities the existence of a bias in the estimation of the exponent as compared to non-linear regression techniques. In spite of this, we choose to use the log-log linear regression for two reasons: first, to facilitate the comparison with other studies, which use this method in their large majority, and second, because maximal velocities in speech are relatively small, essentially below 50 cm/s.

The linear regressions were calculated with the statistical software SPSS for Windows, version 15.0 (Function: Regression, linear model), for each speaker or simulation condition, and each sensor or tongue region separately. For the whole set of data (**All**), a regression was computed for all the samples of the corpus taken together. When data were split in subsets (sentences or segments), regressions were calculated for each subset separately. Only the regressions that were statistically significant ( $p < 0.01$ ) were taken into account for the rest of the study. For the whole set of data (**All**) and for the set **Sent**, all regressions were significant. For the data set **Seg**, a certain percentage of regressions (less than 15% in all cases) was not significant. Table 1 summarizes for each sensor the number of segments selected for each speaker or for each simulation condition, together with their mean path lengths and their mean durations.

----- Insert Table 1 around here -----

The results were further analyzed for the different sets of data (**All**, **Sent**, **Seg**) separately. For the whole set of data (**All**) the analysis focused on two specific aspects: 1.) the exponent values - a - and the velocity gain factor - k - and 2.) the amount of variance ( $R^2$ ) explained by the log-log linear regressions. When data were split in subsets (sentences or segments), the distributions of the exponent values and of the velocity gain factors calculated

for all subsets (i.e. all sentences or all segments) were characterized for each sensor and for each speaker or each simulation condition separately, via their mean values and their variance. The comparison between the two sets of data, **Sent** and **Seg**, was based on these measures.

## RESULTS

### *Complete set of data (All)*

----- Insert Figure 3 around here -----

The distribution of the complete sets of the experimental data was plotted in the  $[\log_{10}(c(t)), \log_{10}(v(t))]$  plane, split by speaker and sensor. The same representation was also done for the simulated data, split by tongue region, and by speaking rate condition. Thus, 35 scatter plots (4 sensors x 5 speakers, 3 sensors for speaker F2, 4 flesh points x 3 speaking rates for the model) were available which summarized the global relations between speed and curvature in the log-log plane. These plots showed that the data distributions were very similar across speakers, languages, and tongue sensors for the experimental data, as well as across tongue regions, and speaking rate conditions for the simulated data. Moreover, strong similarities exist between the experimental and the simulated data. Figure 3 shows two examples of these distributions in the log-log plane, one for the simulated data (top panel) and one for the experimental data (bottom panel). The data are represented in light gray, while the best linear fit is depicted with a bold solid line. The exponent value -  $a$  - and the explained variance ( $R^2$ ) of the linear regression are also indicated in each plot. This figure illustrates well the main trends observed in our data. Both scatter plots look similar. The ranges of variation of speed and curvature are very close, slightly larger though for the simulated data. The exponent values and the explained variance are also very similar. In both cases it can also be observed that the proximity of the data to the linear regression line decreases when curvature becomes very small. This can be easily explained by the inaccuracy of the curvature estimation when the path becomes close to a straight line.

----- Insert Figure 4 around here -----

Figure 4 displays the results of the log-log regressions for the exponent values  $-a$  and for the velocity gain factors  $-k$  respectively, split by experimental and simulated data, by speaker, tongue sensor, tongue region, and speaking rate condition. Table 2 provides for each of the regressions the degrees of freedom (df) and the explained variance ( $R^2$ ). All regressions are statistically significant ( $p < 0.005$ ).

----- Insert Table 2 around here -----

The exponent values are all negative, as shown for one example in Figure 3. Figure 4 represents their absolute values. In the left plots of each panel, a horizontal dotted line marks the position of the  $(-1/3)$  key-value proposed by Viviani and colleagues. Obviously, in all cases the exponent values found in our data are very close to this key-value. The exact range of variation is  $[-0.376; -0.291]$  for the experimental data and  $[-0.377; -0.312]$  for the simulated data. A trend exists for the data simulated at a slow speaking rate to have slightly larger exponent values than the data simulated at normal and fast speaking rate. However, these values are still very close to  $(-1/3)$ . Hence, strong similarities exist for the exponent values between experimental data and simulations carried out with the biomechanical model. As observed by Tasko and Westbury (Tasko and Westbury, 2004; p.76), differences exist for each speaker among the exponent values calculated for the different tongue sensors. However, the direction of variation across sensors is speaker-dependent, and no general trend can be found, except for the fact that the tongue tip sensor tends to correspond to higher exponent values than the other sensors (true for all speakers except C2). Variations among tongue regions are also systematically observed for the simulated data, but the tongue tip does not show the highest exponent values.

The velocity gain factor (see right panels in Figure 4) varies from 3.90 (subject G1, blade sensor) to 11.12 (subject F2, tip sensor) for the experimental data, and from 6.25 (tip region at slow speaking rate) to 11.42 (dorsum region at fast speaking rate). Two points

become evident from these observations: (1) the velocity gain factors computed for the simulated data are within the range of variation measured for the experimental data; and (2) there is a large variability across speakers and the differences between two given speakers are similar for all the tongue sensors. These findings suggest that inter-speaker variability of the velocity gain factor reflects differences in speaking rate among speakers. This assumption is supported by the results obtained for the simulated data, since the velocity gain factor increases when speaking rate increases. However, the range of variability observed for the velocity gain factor across speakers, who were all supposed to speak at normal rate, is much larger than the range of variability observed for the simulated data, for which the speaking rate was explicitly modified.

To further analyze this question, we compared more specifically the results obtained for the different speaking rates of the simulated data, with those obtained for speakers F1 and F2, for whom the corpora used in this study were exactly the same. From speaker F1 to speaker F2 the sentence duration is multiplied by 0.82 (1.01s for speaker F1 and 0.83s for speaker F2), while the ratio for velocity gain factor is between 1.56 and 1.68, depending on the tongue sensor (see Figure 4). For the simulated data, when the duration of the speech sequence is divided by 2 (from the slow to the fast speaking rate), the velocity gain factor is multiplied by a factor ranging from 1.3 to 1.44 (depending on the tongue region). Obviously, the relation between change in speaking rate and change in velocity gain factor is very different. This statement is consistent with the classical observation that speaking rate is not systematically related to the velocity of vocal tract articulators. Indeed, a good amount of studies has shown that speakers may adjust speaking rate either by modifying mainly the overall velocity of articulators or by modifying mainly the distance covered during a segment. In our simulations the first strategy was used, and this is well accounted for by velocity gain factor variations. As shown by sentence durations, speakers F1 and F2 did speak at fairly similar speaking rates, but the kinematic analysis reveals that their articulators moved at

different speeds: for speaker F1 the average velocities of the tongue tip, tongue blade and tongue dorsum were respectively 6.44, 5.24 and 5.25 cm/s, while they were respectively 11, 10.84 and 9.27 cm/s for F2. This corresponds to a speed ratio varying between 1.71 and 2.07, which is also fairly well accounted for by velocity gain factors computed for each speaker. Hence, while the velocity gain factor seems to be a reliable parameter to assess inter-speaker differences in articulators' speed, it is not adapted to measuring inter-speaker differences in speaking rate, because of the variety of strategies available to adjust speaking rate.

The variance explained by the log-log linear regression varies in the range from 0.522 to 0.665 for the experimental data, and in the range from 0.486 to 0.580 for the simulated data. The fit is slightly better for the experimental data. However, in all cases the regressions are statistically highly significant ( $p < 0.005$ ), and given the very large number of degrees of freedom (df, in Table 2) the amount of explained variance can be considered as fairly large.

In summary, the analysis of the complete sets of data, split by speaker, speaking rate, and tongue sensor/region suggests that the 1/3 power law applies to tongue movements. It also shows that the characteristics of the movements generated with the biomechanical tongue model controlled on a target-to-target basis, without any centrally controlled specification concerning the trajectory between the targets, match those of the experimental data well. In addition, our simulations demonstrate that the tongue behavior in the pharyngeal region, which could not be studied experimentally, does not differ from that in the palatal region.

#### *Data split by sentences (**Sent**) and segments (**Seg**)*

It should be acknowledged that the amount of variance explained by the log-log regression in our data (around 50%) is smaller than the amount of variance found in other human movements (see for example deSperati and Viviani, 1997). There are two possible and compatible explanations for this limitation: (1) the velocity gain factor varies among single movements; or (2) the exponent value itself varies among single movements. To further study

this question, log-log linear regressions were computed for the data split according to sentences and segments (see details in section *Data analyses*).

In the experimental data, sentences lasted on average for 1.5 s, while duration of the 3 target sentences of the simulated data comprised between 0.36s (fast speaking rate) and 0.72s (slow speaking rate). Because of this temporal discrepancy, no reliable comparison between simulated and experimental data was possible for the sentences. This is why only data split by segment are presented for the simulations.

----- Insert Figure 5 around here -----

Typical examples of the exponent distributions are given in Figure 5 for the simulated data (tongue blade sensor at normal speaking rate, top panel) and for the experimental data (speaker F1, tongue blade sensor, bottom panel). Looking at these distributions, three general observations can be made: (1) the values are distributed around the  $(-1/3)$  key-value proposed by Viviani and colleagues; (2) in the experimental data the distribution range is wider for the segments (**Seg**) than for the sentences (**Sent**) with more data points for values closer to zero (larger than  $-0.2$ ); (3) for the segments, the width of the distributions of the simulated data is similar to that of the experimental data, but the exponent values tend to be more negative (with minimal values smaller than  $-0.45$ ). Figure 6 (left and central plots of each panel) gives a more detailed account of these general observations.

----- Insert Figure 6 around here -----

The absolute values of the average exponent calculated for all the sentences (**Sent**) (total number of sentences: 1439, with df varying between 8 and 1225) and segments (**Seg**) (total number of segments: 11632, with df varying between 4 and 682) are presented for each speaker or simulation condition, and for each sensor or tongue region separately (left plots). The corresponding standard deviations are presented in the center plots. The results are plotted with light markers for the sentences and with dark markers for the segments. As suggested by Figure 5, it can be observed that all the average values remain in the range of the

(-1/3) key-value. Comparing Figure 4 and Figure 6, it can be noted for the experimental data that the average exponent values calculated for the sentences are very close to the exponent values calculated for the complete set of data (**All**). There is a trend for the average exponents calculated for the segments to have smaller absolute values than the ones calculated for the sentences. For the experimental data the standard deviations calculated for the segments are noticeably larger than those calculated for the sentences. In the majority of cases, a multiplying factor in the range of 2 to 3 is observed between the standard deviations of these two sets of data. For the segments the standard deviation can be 30% of the average value.

----- Insert Table 3 around here -----

Similar observations can be made for the velocity gain factor. The average values presented in the right plots of Figure 6 are close to the velocity gain factors calculated for the complete sets of data (**All**). The standard deviations (given in Table 3) are generally much larger for the segments than for the sentences, with a multiplying factor of around 2 between the two sets of data.

For the sentences, the standard deviations of the exponent values and velocity gain factors distributions are small, in the range of 10 to 15% of their average values. These average values are very close to the values found for the complete sets of data, which describe the global speed-curvature relations of our data.  $R^2$ , the amount of variance explained for the sentences by the log-log linear regressions, varies between 0.224 (df=89) and 0.953 (df=14) with a median value at 0.596. Hence, for the data split by sentences,  $R^2$  values are quite variable, but their median value is in the range of the values calculated for the complete set of data. Hence, the general trend observed for the complete sets of data also applies to each sentence individually. In other words, a power law with an exponent close to (-1/3) accounts for the data at the level of individual sentences as well.

For the segments, such a power law is also found in the majority of cases, as attested by the fact that the average exponent values are close to (-1/3). However, the variability of



the exponent and velocity gain factor values is much larger than for the sentences, which means that for a non-negligible number of segments the speed-curvature relation can not be accounted for by a power law with an exponent value close to  $(-1/3)$ . This is true both for experimental and simulated data, with similar amounts of variability of the exponent and velocity gain factor values within each kind of subset.  $R^2$ , the amount of variance explained for the segments by the log-log linear regressions, varies between 0.104 (df=89) and 0.990 (df=6) with a median value at 0.602. Hence, contrary to our expectations, the amount of explained variance is not larger for the data split by segments than for the data split by sentences.

----- Insert Figure 7 around here -----

Do segments that do not follow the 1/3 power law, have any peculiarities that could explain this phenomenon? To answer this question, segments having exponent values larger than or equal to -0.15 or smaller than or equal to -0.45 were extracted and analyzed with respect to their duration, their average and maximal tangential velocity and the length of their articulatory path. No evidence could be found for any of these characteristics that would make them different from the segments for which the 1/3 power law applies. Figure 7 gives a few examples of these segments (with an exponent value larger than or equal to -0.15 in the two upper rows, and smaller than -0.45 in the two lower rows). It can be seen that they feature a variety of shapes. Some of these segments are heavily curved, while other segments are rather straight. For some segments the curvature is quite constant along the trajectory, while it varies for other segments. In addition to the articulatory path, Figure 7 presents the variation of the tangential velocity along the trajectory with the thickness of the line: the thicker the line, the faster the movement. This presentation allows one to see the detailed relation between curvature and speed along the path. It becomes evident that for all the segments the velocity decreases when curvature increases, which is consistent with a power law with a negative exponent in the form of equation (2), but with an exponent value different

from  $(-1/3)$ . The systematic analysis of their kinematic characteristics revealed that these segments often correspond to small movements with small path lengths and small tangential velocity, especially those having an exponent value larger than  $-0.15$ . However, this is not systematic and many segments with similar small path lengths are well described by the  $1/3$  power law. Hence, it is very unlikely that the inability of the  $1/3$  power law to describe the speed-curvature relation in these segments could be the consequence of an artefact due the small length of the path followed during this segment.

Finally, going back to the question that was raised at the beginning of this section, it seems that the reduced amount of explained variance found in our experimental and simulated speech data, as compared to the values classically published in the literature for other human movements, could have its origin in the variability of the speed curvature relation across individual segments. This variability is found both in the exponent values and in the velocity gain factor values.

## DISCUSSION

On the basis of our results we suggest for speech movements three main conclusions that will be developed below:

- (1) The  $1/3$  power law accounts for general trends of speed-curvature relations; this is true in the experimental data independently of the language considered and in the movements simulated with a biomechanical model of the tongue without any centrally controlled specification of trajectory properties.
- (2) The  $1/3$  power law does not account for the variety of the detailed time-to-time relation between speed and curvature.
- (3) There are numerous results supporting the hypothesis that the global properties accounted for by the  $1/3$  power law do not reflect any property of the underlying motor control strategies.

*General trend of the speed-curvature relations*

In all cases, whatever the data set (**All**, **Sent** or **Seg**), the exponent value is negative, which means that the velocity decreases when the curvature of the articulatory path increases. This is in agreement with the form of Viviani and colleagues' law. In addition, for each subject or modeling condition, and for each tongue sensor or tongue region, the average exponent values calculated across segments are close to the values calculated for the sentences and for the complete sets of data. These values are very close to the (-1/3) key-value proposed by Viviani and colleagues. This is true for the 6 speakers and for the three languages that were analyzed as well as for the simulated data. The simulated data provided evidence that this trend applies also to the pharyngeal region, which could not be experimentally observed because of the limitations of the experimental devices.

With regard to the constant factor of the log-log linear regression, it is shown for the simulated data that it increases with the speaking rate. This supports Viviani and colleagues' hypothesis that this coefficient accounts for intra-speaker variations in the velocity of articulatory movements. However, the large variation of this factor observed across subjects calls into question its reliability for an inter-speaker comparison of speech movement velocities and speaking rate. In summary, if speakers are considered separately, the 1/3 power law indeed accounts for general trends of the speed-curvature relation.

*Time-to-time relations between speed and curvature*

The detailed time-to-time relations between speed and curvature in speech movements can be analyzed on the basis of the log-log linear regressions carried out separately for each individual segment (Figures 5 and 6, and Table 3). Both for experimental and simulated data, the distributions of the slope of the log-log regression, which correspond to the exponent value of the power law relation, show large standard deviations. Obviously, this observation

is not compatible with a law assuming a constant exponent value that applies for each speech movement on a time-to-time basis.

*The 1/3 power law and motor control strategies*

Our results show that the general trends accounted for by the 1/3 power law are observed both for the experimental and the simulated data. It is important to recall here that the simulations carried out with the biomechanical tongue model were controlled on a target-to-target basis without any optimization based on kinematic characteristics of the articulatory paths and without any centrally controlled specification of the trajectory properties or velocity profiles. We have also seen that the 1/3 power law does not systematically apply to each individual speech segment. Taken together, these results indicate that, consistent with the conclusion drawn by Gribble and Ostry (Gribble and Ostry, 1996) for arm movements, the 1/3 power law does not necessarily arise from a specific characteristic of the underlying control strategies as suggested by different authors (Viviani and Flash, 1995; Todorov and Jordan, 1998; Harris and Wolpert, 1998) for other human movements. Thus, the suggestion made by Tasko and Westbury (Tasko and Westbury, 2004, p.77) to use “*the power law and its deviations*” for “*evaluating the success or failure of models that rely on whole trajectory templates [...] or only a few sequential positions*” do not seem to be appropriate.

Since the 1/3 power law cannot be attributed to the underlying motor control strategies of speech, explanations for the law have to be found in intrinsic characteristics of the collected data. From this perspective, the different suggestions (Gribble and Ostry, 1998; Maoz et al., 2005) already mentioned above (Section “*Rationale: the 1/3 power law*”) are particularly interesting. The characteristics of the data simulated with our biomechanical tongue model for speech support Gribble and Ostry's hypothesis that the power law can be explained by the near second order dynamical characteristics of the motor system. Tongue stiffness (i.e. the Young modulus of tongue tissues) varies in the model during the course of a movement due to

muscle activation (see Payan and Perrier, 1997 for details). It could, at least in part, be at the origin of the variability observed in the exponent value across individual segments around the theoretical (-1/3) value. Indentation measurements carried out on a fresh human cadaver tongue have shown that even more complex non-linear relations exist between stress and strain in real tongue tissues (Gérard et al., 2005), which can induce noticeable variations in the stiffness during tongue displacements. Larger variation in the elastic characteristics of the tongue compared with other motor systems could explain why the variability of the exponent values measured in this study for speech is larger than that previously observed for other human movements.

According to Maoz and colleagues (Maoz et al., 2005), noise in the data could largely contribute to the emergence of the 1/3 power law. Their demonstration is based on the formula used to compute the curvature (see equation 4), which, *de facto*, induces that  $\log_{10}(v(t))$  and  $\log_{10}(c(t))$  are linked by the equation:

$$\log_{10}(v(t)) = -\frac{1}{3}\log_{10}(c(t)) + \frac{1}{3}\log_{10}\left(\left|\frac{d^2x}{dt^2}\frac{dy}{dt} - \frac{d^2y}{dt^2}\frac{dx}{dt}\right|\right) \quad (6)$$

If, due to noise, the term  $\frac{1}{3}\log_{10}\left(\left|\frac{d^2x}{dt^2}\frac{dy}{dt} - \frac{d^2y}{dt^2}\frac{dx}{dt}\right|\right)$  is statistically not correlated with  $\log_{10}(v(t))$ , the 1/3 power law directly emerges from equation 6. Given that the accuracy of a correlation estimate between two stochastic processes increases with the number of measurements, the fact that the exponent values estimated from the complete sets of data tend to be closer to (-1/3) than the values estimated from the segments (**Seg**) supports Maoz and colleagues' suggestion.

In conclusion, our study combining the analysis of articulatory data from three different languages and the analysis of articulatory data generated with a biomechanical tongue model provides evidence that the 1/3 power law is a very global characteristic of the speed-curvature relations in speech movements, and that it is language independent. The fact

that it applies to movements simulated with a biomechanical model of the tongue, which was driven without any centrally controlled specification of trajectory properties, suggests that the 1/3 power law does not result from any motor control strategy. In addition, this law does not apply to every single movement description. Hence, no reliable inference can be made regarding speech motor control on the basis of the 1/3 power law.

## **Footnotes**

---

<sup>1</sup> The post-processing procedure was mainly developed by Phil Hoole in Munich (Germany).

## ACKNOWLEDGMENTS

The authors want to express their thanks to Christophe Savariaux and Jörg Dreyer for their technical expertise during the EMMA recordings at ICP and at ZAS, and to the two anonymous reviewers for their very helpful comments. Thanks to our colleagues Christine Mooshammer, Jianwu Dang, and Liang Ma with whom we collected the experimental data for other research purposes, to Laura Koenig for suggestions and proof-reading, to Uwe Biehl for technical assistance, and to Dieter Fuchs.

## GRANTS

This work was supported by the POPAART project funded by the CNRS, the French Ministry of Foreign Affairs in the framework in the P2R program, and the German Research Council (DFG) GWZ 4/8, P1.



REFERENCES

1. **deSperati C, Viviani P.** The relationship between curvature and velocity in two dimensional smooth pursuit eye movement. *The Journal of Neuroscience* 17(10): 3932–3945, 1997.
2. **Feldman AG.** Once more on the Equilibrium-Point Hypothesis ( $\lambda$  Model) for motor control. *J Mot Behav.* 18 (1): 17-54, 1986.
3. **Gérard JM, Ohayon J, Luboz V, Perrier P, Payan, Y.** Non linear elastic properties of the lingual and facial tissues assessed by indentation technique. Application to the biomechanics of speech production. *Medical Engineering and Physics* 27: 884–892, 2005.
4. **Gribble PL, Ostry DJ.** Origins of the power law relation between movement velocity and curvature: modeling the effects of muscle mechanics and limb dynamics. *J Neurophysiol* 76: 2853–2860, 1996.
5. **Harris CM, Wolpert DM.** Signal-dependent noise determines motor planning. *Nature* 394(6695): 780–784, 1998.
6. **Hoole P, Nguyen N.** Electromagnetic articulography in coarticulation research. In: *Coarticulation: Theory, Data and Techniques* edited by Hardcastle WH and Hewlett N. Cambridge, UK: Cambridge University Press, 1999, p. 260-269.
7. **Kandel S, Orliaguet JP, Viviani P.** Perceptual anticipation in handwriting: The role of implicit motor competence. *Perception and Psychophysics* 62(4): 706–716, 2000.
8. **Ladefoged P, Maddieson I.** *The sounds of the world's languages.* Oxford: Blackwell, 1996.
9. **Lacquaniti F, Terzuolo C, Viviani P.** The law relating kinematic and figural aspects of drawing movements. *Acta Psychologica* 54: 115–130, 1983.
10. **Maoz U, Portugaly E, Flash T, Weiss Y.** Noise and the two-thirds power law. *Proceedings of the Neural Information Processing Systems Conference*, 2005.

11. **Ma L, Perrier P, Dang J.** Anticipatory coarticulation in vowel-consonant-vowel sequences: A crosslinguistic study of French and Mandarin speakers. *Proceedings of the 7<sup>th</sup> International Seminar on Speech Production*: Ubatuba, Brazil, 2006.
12. **Massey JT, Lurito JT, Pellizzer G, Georgopoulos A.** Three-dimensional drawings in isometric conditions: relation between geometry and kinematics. *Experimental Brain Research* 88(3): 685–690, 2002.
13. **Mooshammer C, Perrier P, Fuchs S, Geng C, Pape D.** An EMMA and EPG study on token-to-token variability. *Arbeitsberichte des Institutes für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK)* 36: 46-63, 2004.
14. **Payan Y, Perrier P.** Synthesis of V-V Sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication* 22 (2/3): 185-205, 1997.
15. **Perkell J, Cohen M, Svirsky M, Matthies M, Garabieta I, Jackson, M.** Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92: 3078-3096, 1992.
16. **Perrier P, Payan Y, Zandipour M, Perkell J.** Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America* 114(3): 1582-1599, 2003.
17. **Perrier P, Ma L, Payan Y.** Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. *Proceedings of Interspeech*, International Speech Communication Association: 1041-1044, 2005.
18. **Pollick F.E., Ishimura G.** The three-dimensional curvature of straight-ahead movements. *J Mot Behav.* 28(3), 271-279.
19. **Schaal S, Sternad D.** Origins and violations of the 2/3 power law in rhythmic 3D movements. *Experimental Brain Research* 136: 60-72, 2001.

20. **Tasko SR, Westbury JR.** Speed–curvature relations for speech-related articulatory movement. *Journal of Phonetics* 32: 65–80, 2004.
21. **Todorov E, Jordan M.** Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. *J Neurophysiol* 80(2): 696–714, 1998.
22. **Vieilledent S, Kerlirzin Y, Dalbera S, Berthoz A.** Relationship between velocity and curvature of a human locomotor trajectory. *Neuroscience Letters* 305(1): 65–69, 2001.
23. **Viviani P, Terzuolo C.** Trajectory determines movement dynamics. *Neuroscience* 7: 431–437, 1982.
24. **Viviani P, Cenzato M.** Segmentation and coupling in complex movements. *Journal of Experimental Psychology: Human Perception and Performance* 11: 828–845, 1985.
25. **Viviani P, Stucchi N.** Biological movements look uniform: evidence of motor-perceptual interaction. *Journal of Experimental Psychology: Human Perception and Performance* 18: 603–623, 1992.
26. **Viviani P, Flash T.** Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance* 21(1): 32–53, 1995.

## **Table Legendes**

**Table I:** Number of segments (Nseg), mean path length (Path) and mean duration (Dur) for the whole set of segments, split by sensor coil and by speaker for the experimental data, and split by tongue region and speaking rate condition for the simulated data

**Table II:** Explained variance ( $R^2$ ) and degrees of freedom (df), for the complete set of data (**All**), split by sensor coil and by speaker for the experimental data, and split by tongue region and speaking rate condition for the simulated data

**Table III:** Standard deviations of the distribution of the velocity gain factor calculated for the sentences (**Sent**) and the segments (**Seg**) for the experimental and for the segments only for the simulated data.

## **Figure Legends**

**Figure 1:** Examples of data collected with the electromagnetic mid-sagittal magnetometer (AG100). Top panel: schematic representation of the positions of the sensor on the tongue contour in the mid-sagittal plane of the head; from the left to the right: Ttip, Tblade, Tdors, Tback. Bottom panel: trajectories of the Tdors sensor during a 1s long speech sequence for 3 speakers of the three studies languages; from the left to the right: French, Mandarin Chinese, German.

**Figure 2:** Tongue muscle representation in the 2D biomechanical model. The finite element mesh representing the tongue is plotted in light solid lines. The grey filled elements correspond to the anatomical location of the muscles; their stiffness increases when muscle is activated. The solid contours represent the projection of vocal tract contours in the mid-sagittal plane (lips are on the left hand side, pharyngeal walls on the right hand side, glottis is at the bottom on the right). Dark solid lines represent muscle macrofibers. The hyoid bone is presented in dotted lines.

**Figure 3:** Examples of the distribution of the complete set of data (**All**) and of the corresponding best linear fit (dark solid line) in the  $\log_{10}(\text{curvature})$ - $\log_{10}(\text{velocity})$  plane, for the tongue blade sensor/node. Top panel: simulated data. Bottom panel: experimental data, speaker C1.

**Figure 4:** Absolute values of the exponent  $-a$ - (left plots) and velocity gain factors  $-k$ - (right plots) computed for the complete set of data (**All**) for all subjects (C1 & C2: Chinese subjects, plot symbol 'x'; G1 & G2: German subjects, plot symbol 'o'; F1 & F2: French subjects, plot symbol '◇') and for all modeling conditions (Sl: slow speaking rate; No: normal speaking rate;

Fa: fast speaking rate; plot symbol '+'). The horizontal dotted line in the left plots marks the reference position of (-1/3) exponent value. From top to bottom: tongue back sensor (experimental data) and pharyngeal node (simulated data); tongue dorsum sensor/node; tongue blade sensor/node; tongue tip sensor/node.

**Figure 5:** Examples of the measured probability distribution of the exponent values -a-. Top panel: simulated data (blade node, normal speaking rate), distribution for the set of segments (**Seg**) (number of segments: **624**). Bottom panel: experimental data (speaker F1, blade sensor), distribution for the set of segments (**Seg**) (number of segments: **138**) (solid line) and for the set of sentences (**Sent**) (number of sentences: **50**) (dashed line).

**Figure 6:** Absolute values of the exponent -a- (left plots), standard deviations of the exponent (central plots) and velocity gain factors -k- (right plots) computed for the data split by sentences (**Sent**) (light symbols) and by segments (**Seg**) (dark symbols). See figure 4 for additional details.

**Figure 7:** Some examples of trajectories collected for segments of the experimental data having an exponent value -a- larger than -0.15 (the two upper rows) or smaller than -0.45 (the two lower rows). The thickness of the trajectory is proportional to a quantified representation of the tangential velocity: the thicker the line, the larger the tangential velocity.

Table 1. Number of segments (*Nseg*), mean path length (*Path*) and mean duration (*Dur*) for the whole set of segments, split by sensor coil and by speaker for the experimental data, and split by tongue region and speaking rate condition for the simulated data

Speaker	Tongue tip			Tongue blade			Tongue dorsum			Tongue back/pharynx		
	Nseg	Path (cm)	Dur (ms)	Nseg	Path (cm)	Dur (ms)	Nseg	Path (cm)	Dur (ms)	Nseg	Path (cm)	Dur (ms)
<i>Experimental data</i>												
C1	270	2.98	411	243	2.97	460	263	2.86	422	269	2.65	417
C2	114	2.29	392	98	3.36	464	91	3.3	500	87	3.36	469
G1	207	1.02	247	198	0.855	250	186	1.01	277	172	1.04	286
G2	176	1.28	292	150	1.38	350	164	1.33	320	165	1.51	319
F1	186	1.71	261	138	1.86	349	116	2.24	419	122	1.86	403
F2	98	4.70	425	98	4.63	427	92	4.22	455			
<i>Simulated data</i>												
Slow	774	1.4	249	824	1.5	237	786	1.90	255	832	1.51	242
Normal	606	1.57	246	624	1.78	244	621	2.22	249	659	1.74	238
Fast	529	1.26	180	542	1.46	183	560	1.87	182	562	1.54	183

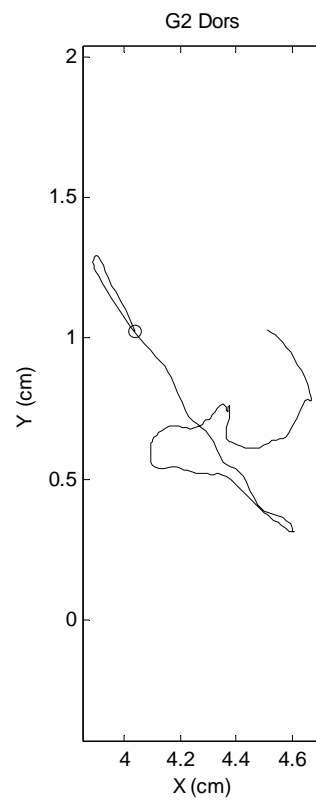
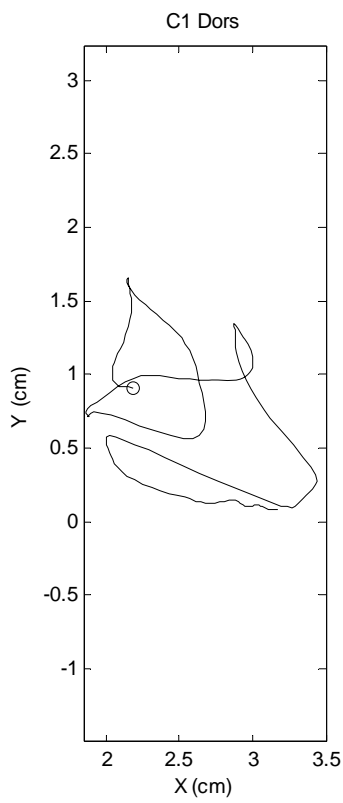
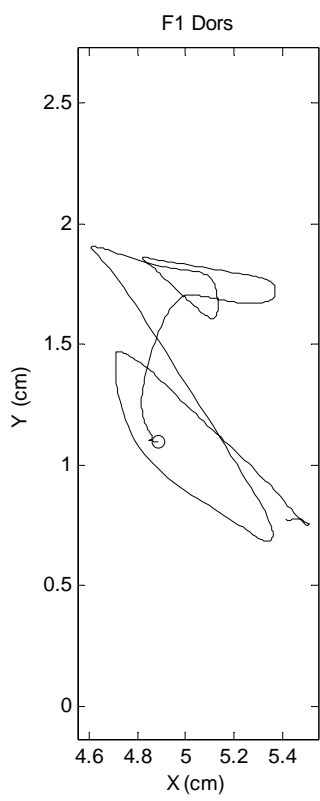
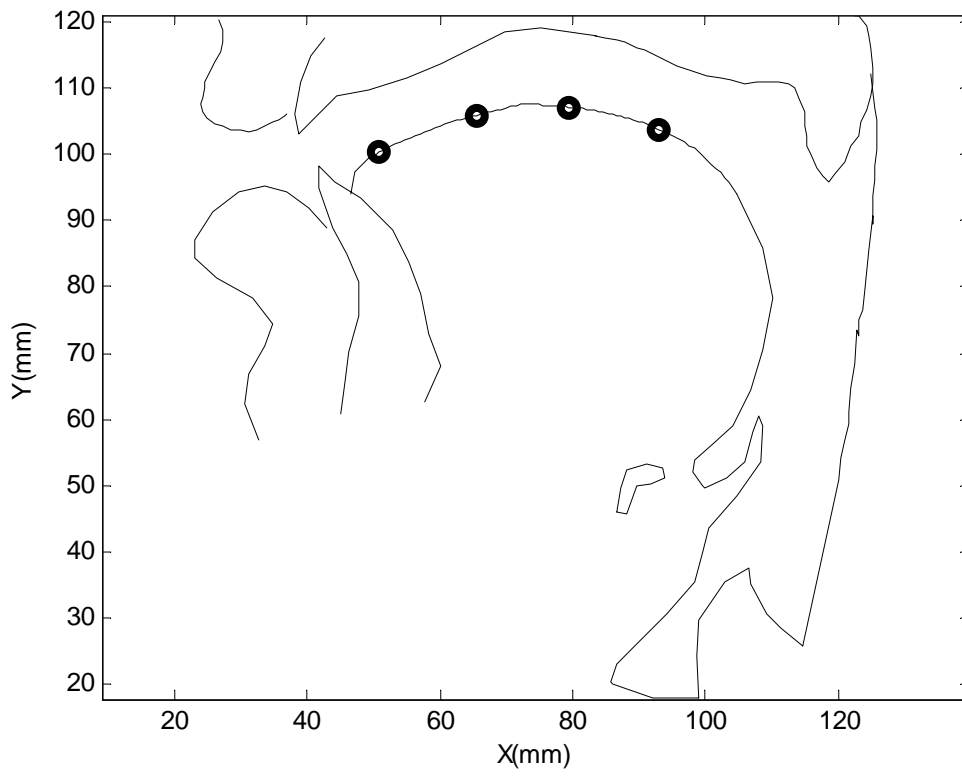
Table 2. Explained variance ( $R^2$ ) and degrees of freedom ( $df$ ), for the complete set of data (*All*), split by sensor coil and by speaker for the experimental data, and split by tongue region and speaking rate condition for the simulated data

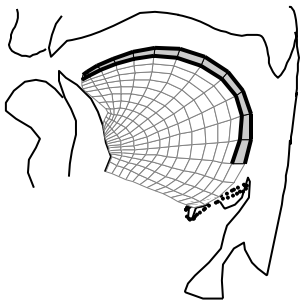
Speaker	Tongue tip		Tongue blade		Tongue dorsum		Tongue. back/pharynx	
	$R^2$	df	$R^2$	df	$R^2$	df	$R^2$	df
<i>Experimental data</i>								
C1	0.596	22831	0.610	23070	0.611	22946	0.639	23011
C2	0.616	10755	0.597	10827	0.607	10817	0.588	10831
G1	0.626	10949	0.536	11109	0.550	11077	0.522	10883
G2	0.559	10828	0.530	11042	0.563	10948	0.607	10938
F1	0.665	24691	0.614	24671	0.593	24847	0.582	24874
F2	0.601	20923	0.605	20950	0.560	20948		
<i>Simulated data</i>								
Slow	0.538	26937	0.580	26810	0.573	26881	0.552	26865
Normal	0.512	20373	0.564	20384	0.540	20207	0.542	20310
Fast	0.486	14101	0.521	14126	0.520	13983	0.531	14071



Table 3. Standard deviations of the distribution of the velocity gain factor calculated for the sentences (**Sent**) and the segments (**Seg**) for the experimental and for the segments only for the simulated data.

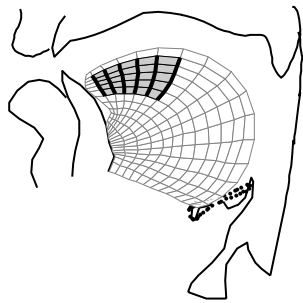
Speaker	Tongue tip		Tongue blade		Tongue dorsum		Tongue back/pharynx	
	Seg	Sent	Seg	Sent	Seg	Sent	Seg	Sent
<i>Experimental data</i>								
C1	2.56	1.05	1.95	0.81	0.611	1.84	1.68	0.88
C2	1.97	1.01	1.83	1.02	1.79	0.913	1.87	0.87
G1	1.01	0.50	1.25	0.51	1.31	0.54	1.47	0.47
G2	1.74	0.85	1.31	0.88	1.54	0.94	1.47	0.71
F1	1.47	1.00	1.44	0.77	1.56	1.00	1.91	1.54
F2	2.66	1.45	2.27	1.12	2.03	0.95		
<i>Simulated data</i>								
Slow	2.31		2.42		2.59		2.44	
Normal	2.25		2.22		2.88		2.37	
Fast	3.04		2.63		3.88		2.72	





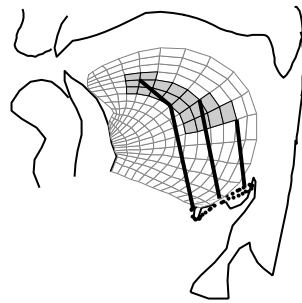
Superior Longitudinal

+



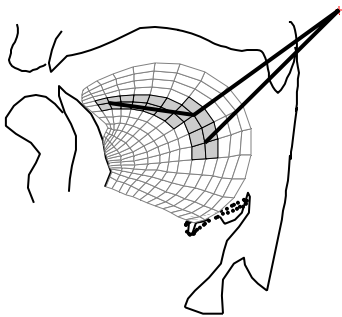
Verticalis

+

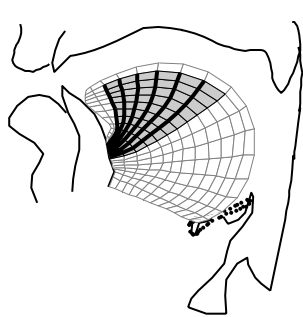


Hyoglossus

+

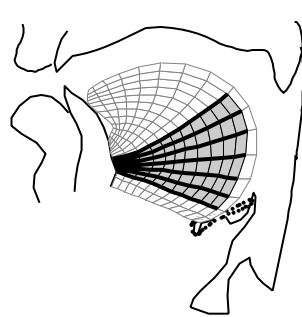


Styloglossus



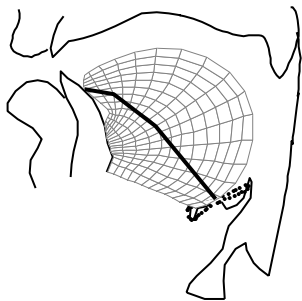
Anterior Genioglossus

+



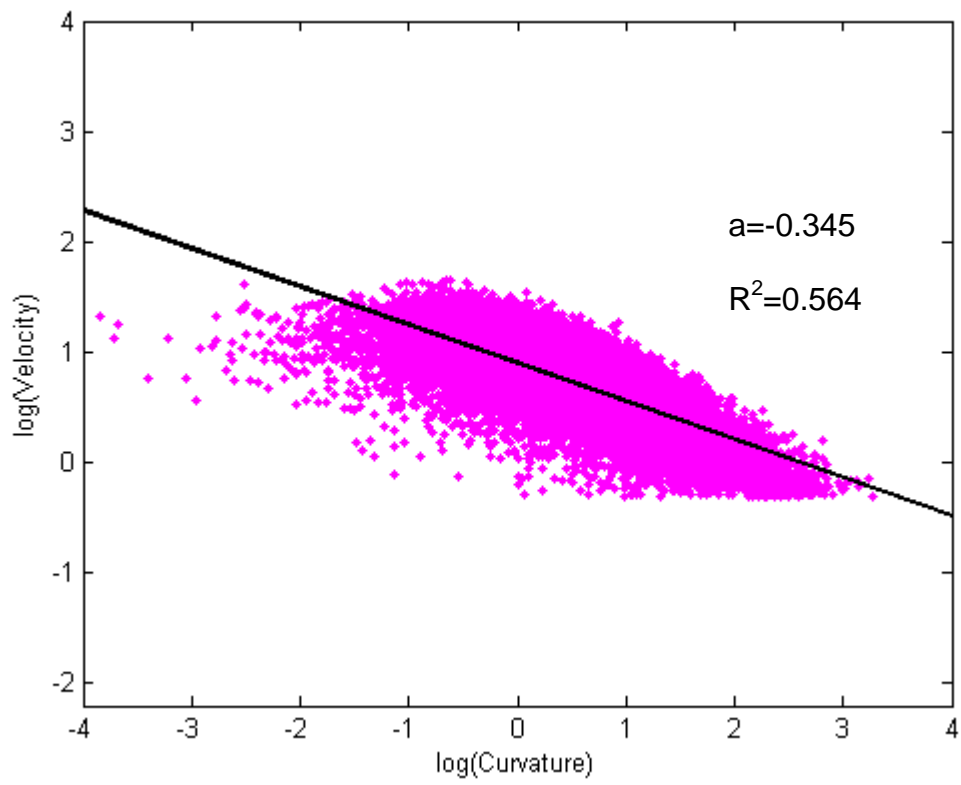
Posterior Genioglossus

+

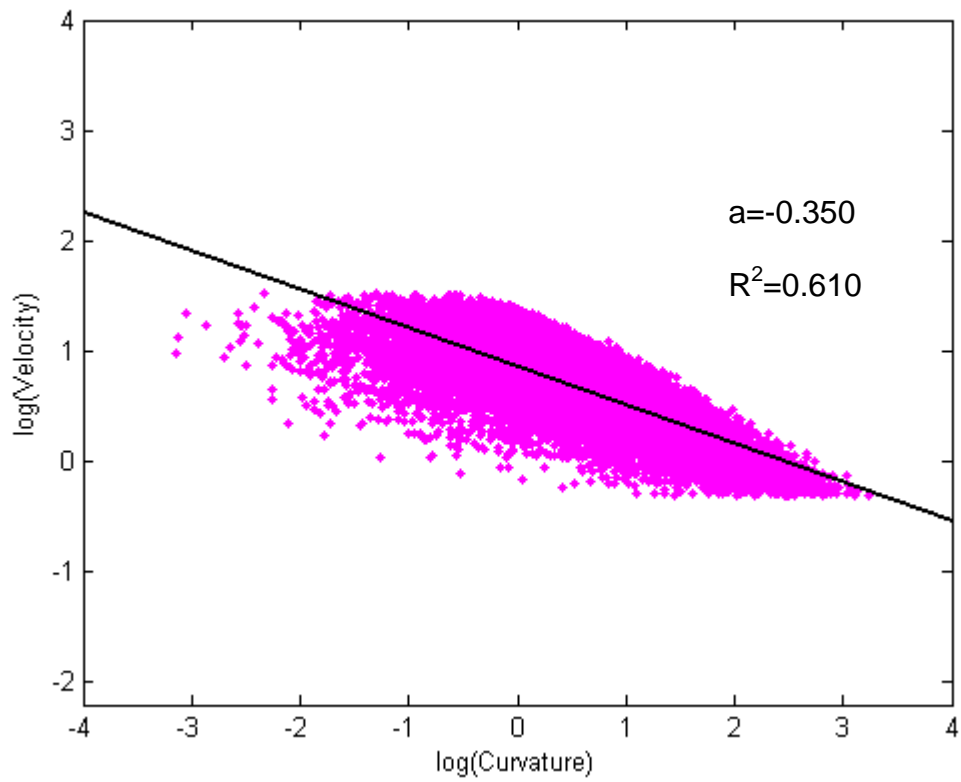


Inferior Longitudinal

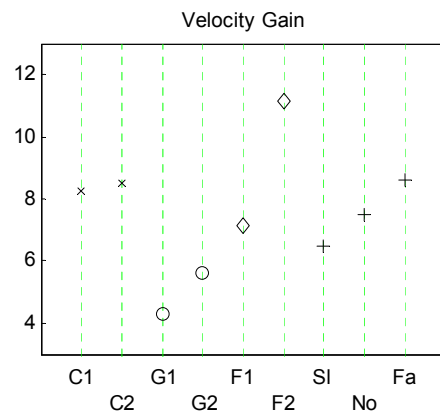
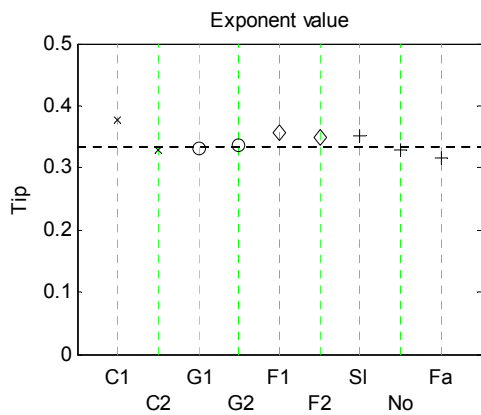
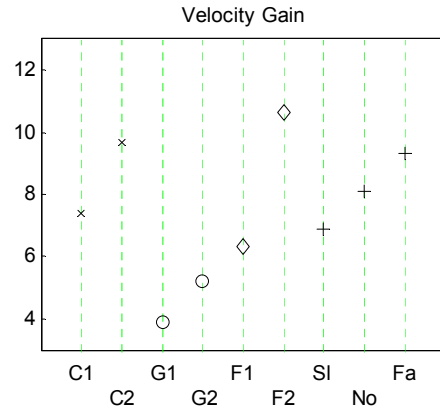
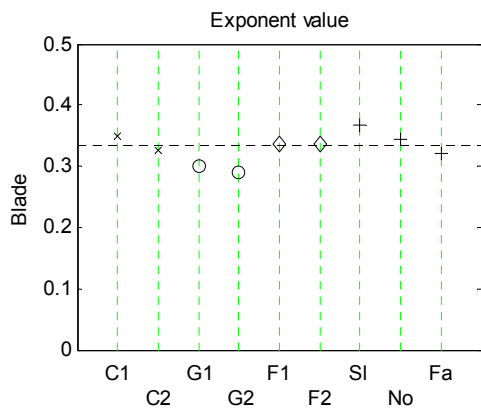
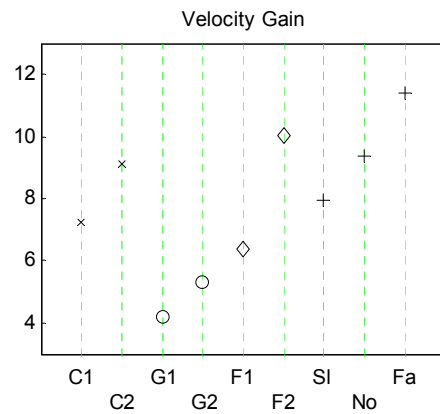
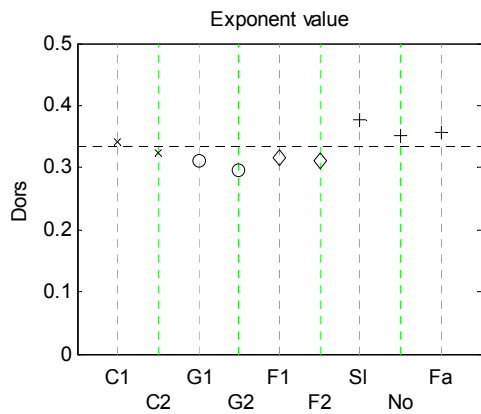
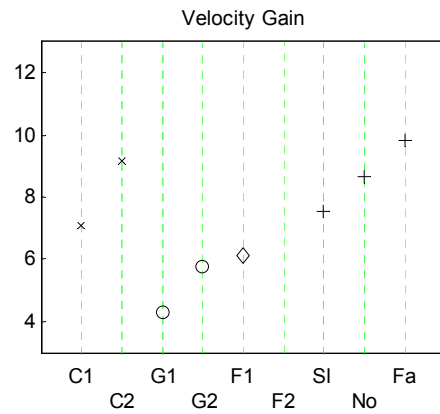
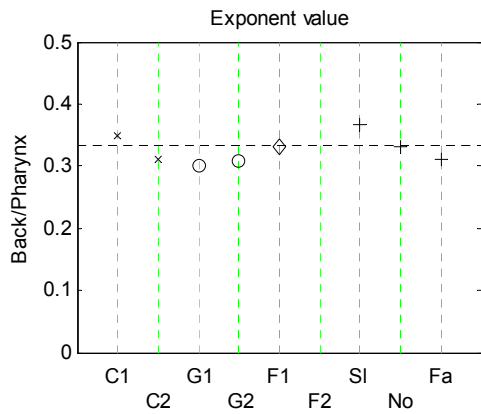
+

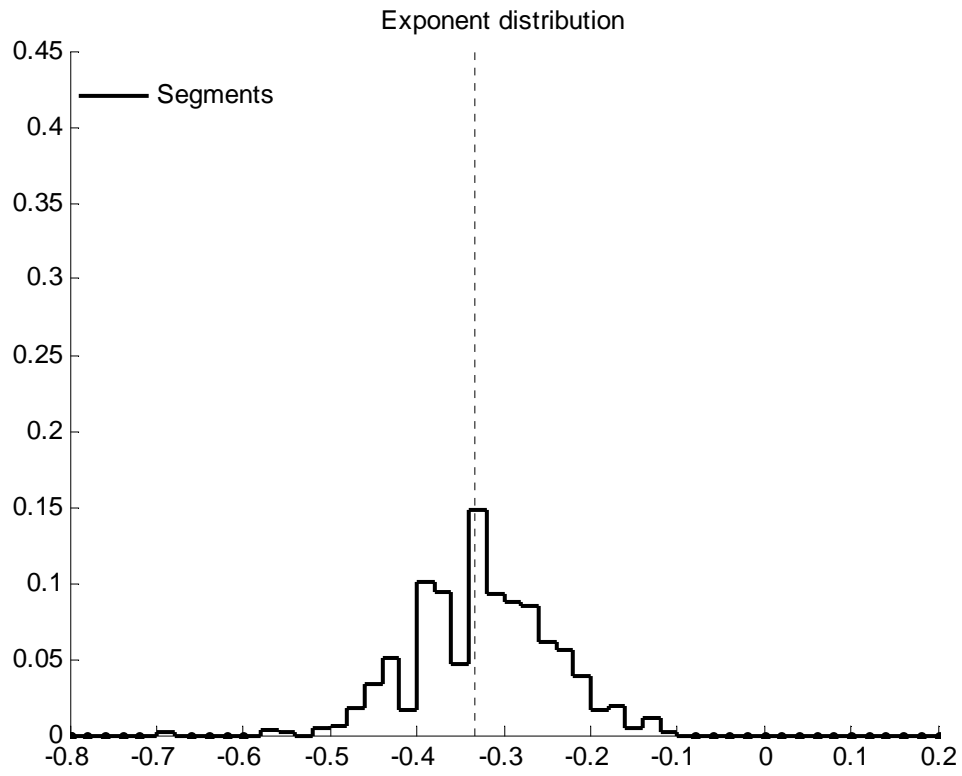


**(a) Blade – Model**

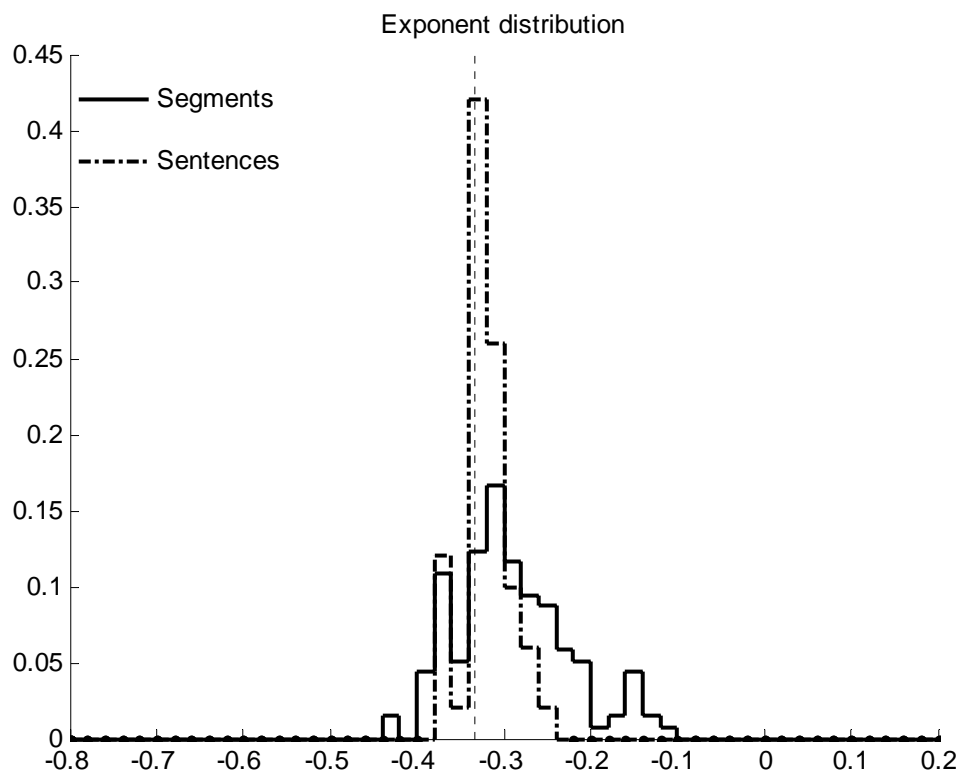


**(b) Blade – Speaker C1**

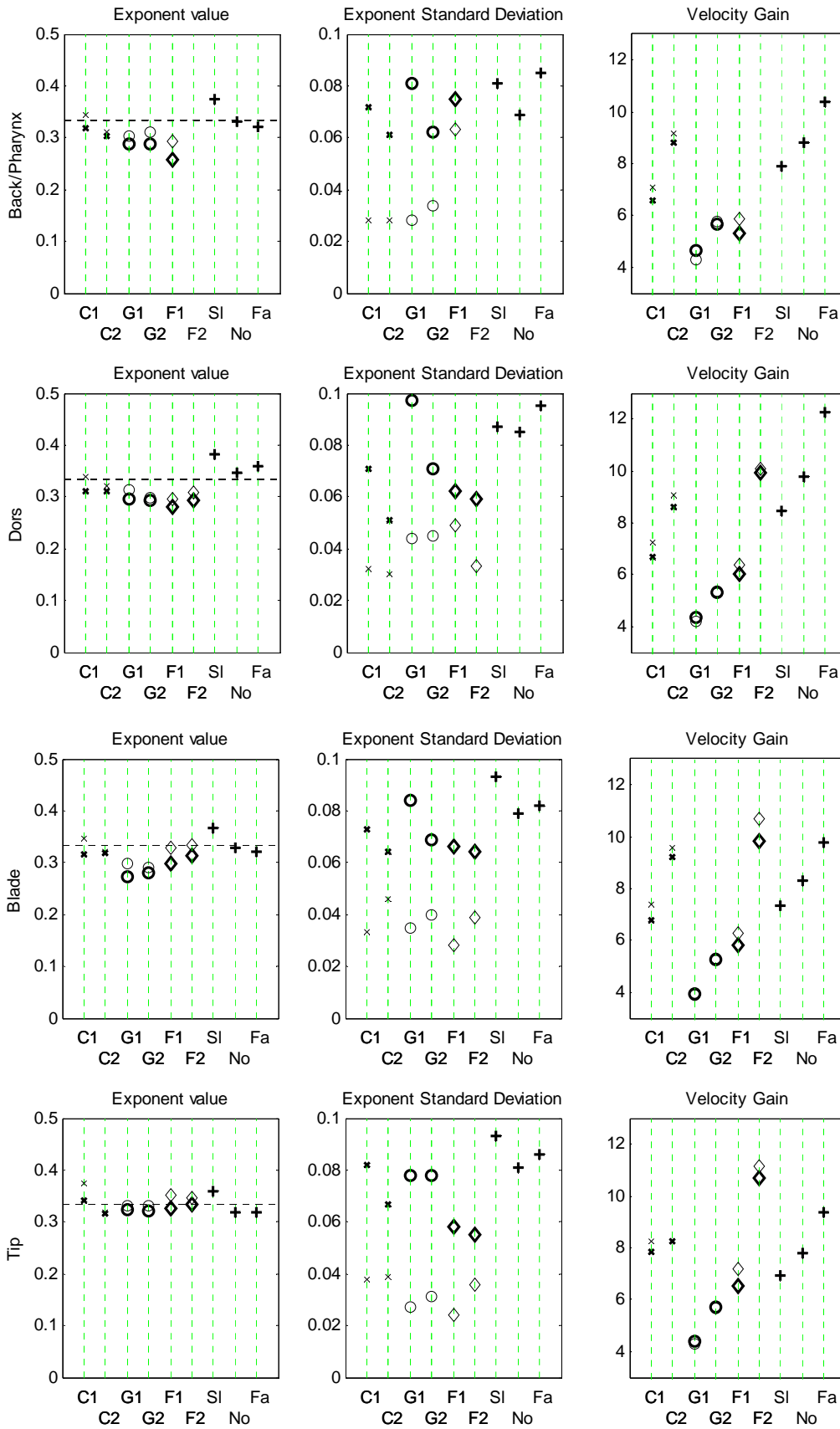




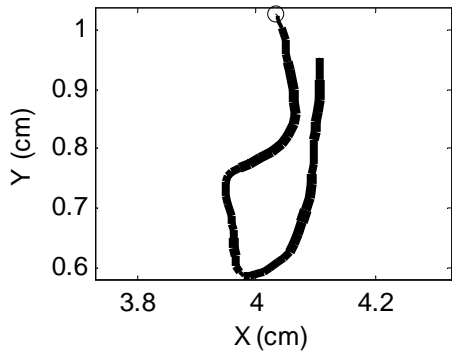
**(a) Blade – Model**



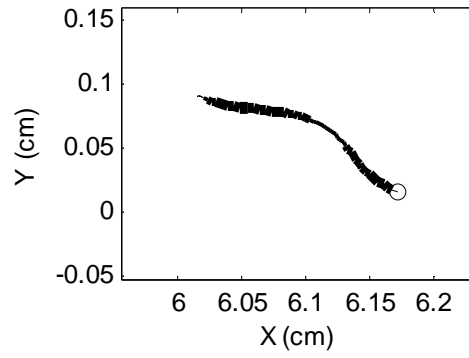
**(b) Blade – Speaker F1**



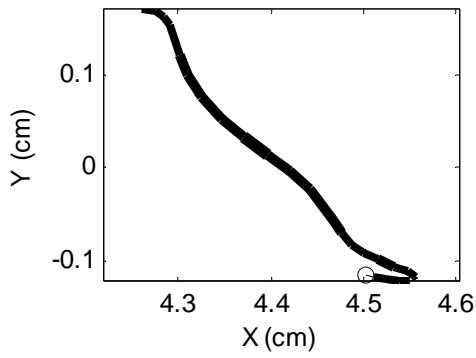
G1 - Dors - Exponent=-0.10



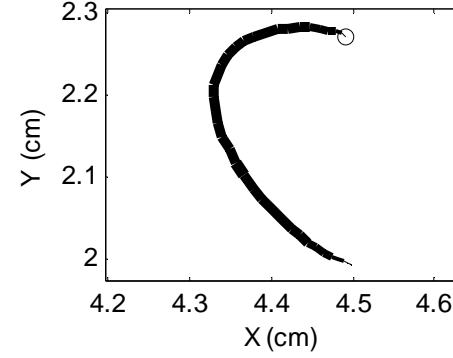
F1 - Back - Exponent=-0.09



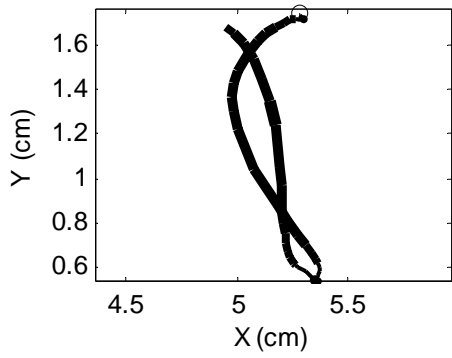
C1 - Back - Exponent=-0.11



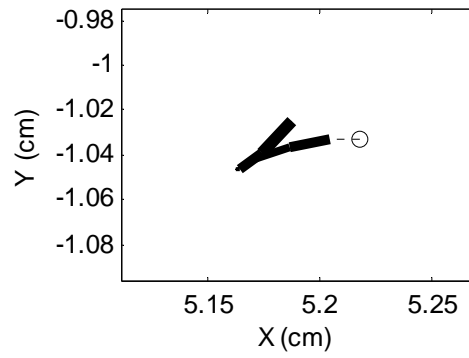
F2 - Dors - Exponent=-0.15



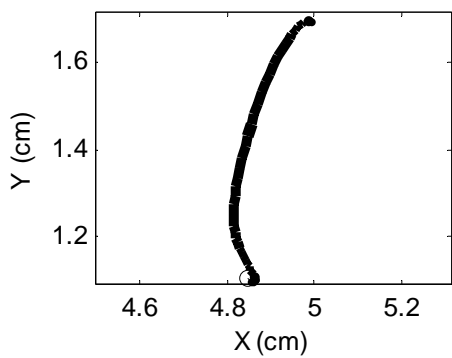
G2 - Back - Exponent=-0.48



C2 - Back - Exponent=-0.52



F1 - Dors - Exponent=-0.48



G1 - Dors - Exponent=-0.46

