

A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model

Susanne Fuchs

Center for General Linguistics (ZAS/Phonetik), Berlin, Germany

Pascal Perrier

DPC/Gipsa-lab, CNRS, Grenoble INP, Grenoble, France

Mariam Hartinger

Center for General Linguistics (ZAS/Phonetik), Berlin, Germany

Corresponding author:

Pascal Perrier

DPC/Gipsa-lab

961 Rue de la Houille Blanche

BP46

38402 Grenoble Cédex France

Email: Pascal.Perrier@gipsa-lab.grenoble-inp.fr

Abstract

Purpose: Linear second-order models have often been used to investigate properties of speech production. However, these models are inaccurate approximations of the speech apparatus. This study aims at assessing how reliably stiffness can be estimated from kinematics with these models.

Methods: Articulatory movements were collected for 9 speakers of German during the production of reiterant /CVCV/ words at varying speech rates. Velocity peaks, movement amplitudes and gesture durations were measured. In the context of an undamped model, two stiffness estimations were compared which should theoretically yield the same result. In the context of a damped model, gestural stiffness and damping were calculated for each gesture.

Results: Numerous cases were found where stiffness estimations based on the undamped model contradicted each other. Less than 80% of the data were found to be compatible with the properties of the damped model. Stiffness tends to decrease with gestural duration. However, it is associated with a large, unrealistic damping dispersion, making stiffness estimations from kinematic data to a large extent unreliable.

Conclusion: Any conclusions about speech control based on stiffness estimations using linear second-order models should therefore be considered with caution.

Introduction

The modeling of the dynamics of vocal tract articulators by means of a linear second order model (henceforth LSOM) has been very useful in understanding and characterizing some general aspects of speech production. For instance, vowel deletion in fast speech could be predicted by introducing the hypothesis of gestural overlap (Browman & Goldstein, 1990a), which involves competition between second-order dynamical attractors. The relation between articulators' dynamics and gesture durations has been investigated from the perspective of the relations between stiffness and oscillation frequency in the context of an LSOM (Ostry & Munhall, 1985; Kelso et al., 1985). It could be shown that within a gesture, peak velocity V_{max} , movement amplitude Amp , and duration T cannot be treated as independent variables. Because of these and other ambitious studies in this framework (e.g. Saltzman, 1986; Saltzman & Munhall 1989), the LSOM has become very popular in speech production research. However, the model has not only been used to describe general aspects of speech production. It has also been used as a tool to quantitatively infer physical characteristics of articulators, such as stiffness, from speech kinematics. Based on such an approach Kelso et al. (1985) proposed that stiffness and rest position are the key control parameters for speech production and determine speech rate variations. Ackermann et al. (1995) suggest that stiffness can be a key parameter to differentiate speech motor control between populations with different speech disorders. More recently, Perkell et al. (2002) measured stiffness based on undamped LSOM properties and discussed their results in the context of economy of effort in speech production. Kühnert & Hoole (2004) relate alveolar and velar articulatory reductions with respect to stiffness in the context of an undamped model. Similarly, in van Lieshout et al. (2007) stiffness was analyzed to account for control mechanisms in apraxia of speech. Xu & Wang (2009) also proposed using the properties of a damped model to estimate stiffness with respect to the timing of f_0 modulations in Mandarin Chinese. However, there is growing evidence that stiffness cannot be seen solely in the framework of an LSOM, but is

physical in nature and therefore crucially dependent on biomechanics, muscular forces, and cocontraction, among other factors (e.g. Shiller et al., 2005, Gomi et al., 2002).

In the context of a linear second-order model, articulatory stiffness is computed based on the assumption (see Appendix) that the ratio V_{max}/Amp is proportional to the square root of the mass-normalized stiffness K/m :

$$\frac{V_{max}}{Amp} = \alpha \cdot \sqrt{K/m} \quad (1)$$

In addition, articulatory stiffness can also be calculated based on the assumption (see Appendix) that the mass-normalized stiffness K/m is proportional to the duration T of the gesture:

$$\sqrt{K/m} = \beta/T \quad (2)$$

This approach is implicitly based on three assumptions. (a) Speech gestures correspond to half a period of an LSOM oscillating around its rest position. The position reached at the end of the movement should correspond to a turning point (the extreme position of the oscillation). (b) The proportionality factors α and β used in the equations (1) and (2) are assumed to be constant across gestures. (c) Stiffness and damping are constant during the whole duration of a gesture. In speech production, however, these three assumptions are problematic for the following reasons: There are speech sounds, like long vowels, sibilants or geminates, which can be kept relatively stable over fairly long time intervals. During these intervals plateaus are observed in the articulatory trajectory. These plateaus are not compatible with the turning point conception of assumption (a), unless the damping is stronger than the critical damping. However, in this case the influence of damping on peak velocity and gesture duration would be at least as strong as the influence of stiffness. Assumption (b) is only true if the system is undamped or if the same relation between damping and stiffness is found in all gestures (see Appendix). These two conditions are far from being realistic. First, the vocal tract is such a

narrow space that during the course of their displacements, speech articulators are likely to be in contact with various structures (e.g., contact of the tongue with the teeth or the palate). Hence, damping clearly exists due to interaction of the tongue with vocal tract structures and it is likely to vary across gestures as well as over the course of a gesture. Second, the mechanical properties of vocal tract articulators are very different. Tongue, lips, and velum are soft bodies while the jaw is a rigid body. Hence the nature of the damping is not the same for all these articulators. Assumption (c) may not hold true since damping can vary along the course of the articulator's displacement. The same holds true for stiffness. Oro-facial muscles and, more generally, human soft tissue have been shown to have non-linear passive elastic properties for which stiffness (Young Modulus) increases with strain (Fung, 1981; Gerard et al., 2005). Moreover, stiffness increases with the level of muscular activity (McMahon, 1984). Assumption (c) also implies that the control parameters of the LSOM (stiffness and rest position) are set up instantaneously at the beginning of each gesture, without any transition period. Instantaneous changes are indeed non-physiological. Hence, in spite of its strengths in globally describing important trends of speech gestures, the LSOM remains a rough approximation of the physical characteristics of articulatory gestures.

Some of these limitations have already been acknowledged in previous studies. For example, Browman & Goldstein (1985) consider the problem raised by the existence of articulatory plateaus. They realize that plateaus cannot be accounted for without considering a specific time variation in the damping. Browman & Goldstein (1990b) also insist on the necessity of properly evaluating the damping for each gesture, and they agree with proposals made in the literature that the control parameters of the model should not vary as a step function. Kelso et al. (1985) address the fact that stiffness varies as a function of strain. They propose an alternative way to measure stiffness in the context of an undamped model.

The aim of the current study is to assess how reliably gestural stiffness can be inferred from articulatory amplitude and velocity or gestural duration as proposed by the LSOM.

Method

Articulatory data for various consonants, speakers and speaking rates were used to evaluate gestural stiffness in the context of the LSOM. This assessment was carried out first using an undamped linear second-order model and then using a damped LSOM.

Corpus

Subjects were asked to produce reiterant speech first at an increasing and then at a decreasing speaking rate over a time interval of 16 seconds, following an experimental scenario proposed by Rochet-Capellan & Schwartz (2007). The corpus consisted of repetitions of bisyllabic CVCV words chosen from among /fata/, /kata/, /kapa/, /kaʃa/, /pasa/, /paʃa/, /pata/, /paka/, /sapa/, /ʃapa/, /ʃaka/, /tafa/, /tapa/ and /taka/. The variety of consonants and articulators involved and the different speech rates allowed us to evaluate the reliability of the LSOM in stiffness estimation. Movements were recorded by electromagnetic articulography (Carstens, AG 100). Three sensors were glued on the tongue at around 1 cm, 3 cm and 5 cm from the tongue tip. Lower lip and jaw movements were measured using sensors glued at the vermillion border and just below the lower incisors. Reference sensors glued on the upper incisors and on the bridge of the nose were used to compensate for head movements in the helmet. A bite plane served as a reference to determine the vertical and horizontal coordinates of each individual vocal tract. The contribution of the jaw was not subtracted from tongue and lower lip movements, since this decomposition is not straightforward. Movements were sampled at 200 Hz. Movements and velocities were filtered with an 18 Hz and a 15 Hz low-pass Kaiser filter, respectively.

Nine speakers of German were recorded, 5 females (sp1 to sp5) and 4 males (sp6 to sp9). They had no reported history of speech, language or hearing impairment. All procedures were in agreement with our institutional regulations. All subjects provided informed consent prior to testing. Subjects were between 30 and 42 years old, except sp7, who was 68 years old. Sp7

was included in the corpus because he is an ultra-fast speaker who has won several competitions for his fast and precise speech in the mass media (Jannedy et al., 2010). We expected that this speaker might use different articulatory strategies than the other subjects.

The speaking pace was given by a visual metronome (see Rochet-Capellan & Schwartz 2007) consisting of successively varying black and white images displayed on a computer screen. Their frequency of change varied between 3.3 and 20 Hz. Sp7, however, was free to choose the pace range according to his specific skills.

Measures

Closing and opening gestures were analyzed in the present study. They were labeled on the tangential velocity signal using the velocity minimum of the movements toward and from the consonants. To keep the amount of data at a reasonable level for each consonant, only the articulators that are known to be crucial for the production of a given consonant were considered. Tongue tip and jaw movements were analyzed for /t, s/, tongue dorsum and jaw movements for /ʃ/, tongue back movements for /k/, and lower lip and jaw movements for /p, f/. Within each of the segmented gestures and for each of the crucial articulators, movement amplitude Amp , velocity peak V_{max} and gesture duration T were measured. Movement amplitude was computed as the length of the articulatory path by computing the cumulative Euclidian distances from one sample to the next.

Stiffness estimation in the context of an undamped linear second-order model

In an undamped LSOM, equations (1) and (2), computed respectively with $\alpha=0.5$ and $\beta=\pi$ (see Appendix), are theoretically two strictly equivalent ways to compute the mass-normalized stiffness k . Equation (1) is based on kinematic measures, and equation (2) on temporal measures.

Undamped LSOMs have essentially been used in speech production studies to evaluate differences in stiffness between two gestures, and not to consider absolute stiffness values.

For example, a stiffness increase from one gesture to another could be interpreted as an increase in articulatory effort (Perkell et al., 2002). Consequently our study focused on stiffness differences between gestures. The following procedure was carried out: For a given speaker and a given condition (split by word, syllable, consonant and articulator) all possible combinations of two gestures were selected and grouped in pairs. For each pair the stiffness difference was computed, once on the basis of equation 1 and once on the basis of equation 2. All cases were considered as misleading (errors) where differences in stiffness showed a positive (or negative) sign on the basis of one equation and a negative (or positive) sign on the basis of the other. The percentage of gesture pairs for which such opposite estimations (henceforth *OpposEstim*) were found was taken as a measure of the inadequacy of an undamped LSOM to calculate stiffness. Whenever differences in stiffness showed the same sign, either positive or negative, the model was considered to be adequate.

In order to be as conservative as possible, only the following cases were selected: Since movement data were sampled at 200 Hz, a maximal potential inaccuracy of 5 ms exists in the time labeling. Hence, only gestures with a temporal difference longer than 5 ms were taken into account. *OpposEstim* was computed only for pairs in which stiffness differences could not be attributed to measurement inaccuracies. In line with classical principles used for measurement techniques in physics, only cases where stiffness differences correspond to a minimum of 10% of the stiffness for at least one gesture of the gestures pair were taken into account. This enabled the possible consequences of these inaccuracies in our evaluation to be minimized. The distribution of *OpposEstim* was analyzed for each speaker and condition as a function of the duration T_1 and T_2 of the paired gestures. To this end, the gesture duration plane (T_1, T_2) was divided into elementary squares, each 20 ms x 20 ms in size (Figure 1).

Stiffness estimations in the context of a damped second-order model

The second step of the evaluation consisted in finding the mass-normalized stiffness and damping factor in the context of an under-damped LSOM that would account for the

experimental measures of V_{max}/Amp and T . In the context of such a model, the movement of the speech articulator is assumed to obey an equation of the type (3)

$$a(t) + b \cdot v(t) + k \cdot x(t) = 0 \quad (3)$$

where $x(t)$, $v(t)$ and $a(t)$ are the position, the velocity, and the acceleration of the articulator, respectively, and b and k are the mass-normalized damping factor and stiffness (hereafter simply damping and stiffness). The following two equations describe the relation between movement amplitude Amp , velocity peak V_{max} , gesture duration T , stiffness k and damping b (see Appendix for details):

$$T = \frac{2 \cdot \pi}{\sqrt{4 \cdot k - b^2}} \quad (4)$$

and

$$\frac{V_{max}}{Amp} = \frac{\sqrt{k}}{1 + e^{-\frac{b \cdot T}{2}}} \cdot e^{-b \cdot \frac{T}{2} \left[\frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left(\frac{b \cdot T}{2 \cdot \pi} \right) \right]} \quad (5)$$

If no solution can be found for the two equations (4, 5) when the kinematic parameters are experimentally measured for a particular speech gesture, then the production of this gesture cannot be assumed to behave like a damped linear second-order model. In this case gestural stiffness cannot be inferred from speech kinematics. If a solution for stiffness and damping exists for these two equations, an LSOM can be assumed to underlie the production of this gesture. To solve equations (4) and (5) the *fsolve* function in Matlab was used.

Results

Stiffness estimations based on an undamped linear second-order model: Opposite estimations

Considering all gestures together, the mass-normalized stiffness values computed either from equation (1) or from equation (2) ranged from 80 s^{-2} to 11000 s^{-2} . Figure 1 shows the corresponding distribution of the percentage *OpposEstim*. It can be seen that the percentage of opposite estimations is far from negligible (Minimum: 2.9%, Maximum: 47.1%; see Table 1). However, opposite estimations do not occur everywhere in the (T_1, T_2) plane. They are found

in 52.2% of the elementary squares (Table 1). These squares are mostly located in the neighborhood of the diagonal in the (T_1, T_2) plane. No case of opposite estimations was found when a very short gesture was compared with a very long one. The percentage of errors (*OpposEstim*) increases with gestural duration.

When data are split across speakers, similar observations can be made. Differences across subjects are only found in the percentage of elementary squares in which opposite estimations are found. This varies from 33.3% (sp6) to 60% (sp9). No significant difference was found across speakers in the minimum and maximum values of *OpposEstim*. In sum, the trends observed for the distribution of *OpposEstim* for all data taken together are confirmed for each speaker.

Results obtained for a damped linear second-order model

Considering all speakers and all repetitions together, the total number of analyzed gestures is 27531. For 5926 of these gestures, it was not possible to find a stiffness k and a damping b for the damped model that are compatible with both the measured V_{max}/Amp value and the T duration value as specified in equations (4) and (5). This result represents 21.5% of the total number of data. When considering the data of each speaker separately, a similar percentage is found for 7 speakers (Table 2a). For the female speaker sp4, the percentage is clearly smaller (13.3%), and it is clearly larger (29.5%) for the ultra-fast male speaker sp7. When considering the data of all speakers together split by articulators and by gestures (Table 2b), a larger variability is observed: the percentage is significantly smaller than 21.5% in the closing gestures of the tongue tip and the lower lip, and it is significantly larger (close to 40%) for the tongue back. Nevertheless, the proportion of these data is generally substantial, and the value 20% is quite representative of the general trend observed across speakers, articulators and gestures.

To further investigate the characteristics of these 20% of data, all measures from all speakers were plotted in the $(T, V_{max}/Amp)$ plane. For the sake of clarity, it should be noted that for an LSOM, V_{max}/Amp and T are theoretically linked by the relation

$$\frac{V_{max}}{Amp} = \frac{c}{T} \quad (6)$$

where c increases with the damping, with a minimum value equal to 1.57 ($\pi/2$) in the case of an undamped system. Speech data have often been found in the literature to be generally well accounted for by equation (6) with c values larger than 1.57 (see for example Ostry & Munhall (1985), with c values ranging from 1.83 to 1.9), which is consistent with the hypothesis of the properties of an under-damped LSOM.

Figure 2 shows the distribution of the data. Data for which solutions do not exist for equations (4) and (5) (Set1, henceforth) are plotted in light gray. Other data (Set2) are plotted in dark gray. The bold solid curve corresponds to equation (6) with $c=1.57$. All the data of Set1 are located below this curve. Thus, all data of Set1 correspond to c values that are smaller than 1.57, which is not compatible with the characteristics of an LSOM. These data are well accounted for by equation (6) with $c=1.44$ (dashed-dotted curve). All the data of Set2 are above the bold solid line, and they are well accounted for by equation (6) with $c=1.85$ (dotted curve). Table 2 confirms that the properties of the Set1 and Set2 distributions and the c values computed for all data together are well representative of the data split by speakers, articulators and gestures.

In summary, for a significant percentage of the data (around 20%) the kinematic measures V_{max} and Amp and the measured gesture duration T are not related in a way that is compatible with an LSOM. For the data, the ratio V_{max}/Amp is smaller than the smallest value observed for an LSOM having the same gesture duration. These data can be seen as intermediate between gestures generated by an undamped LSOM, in which duration and V_{max}/Amp are fully determined by the stiffness, and constant velocity gestures (corresponding to $c=1$), in

which these variables are fully specified by an extrinsic controller. For these data, no reliable information about gestural stiffness can be extracted either from V_{max}/Amp or from T .

In the rest of this section, only the data of Set2 will be considered. The c values are in good agreement with those found in former studies (for example Ostry & Munhall, 1985). In addition, as predicted by an under-damped LSOM (equation (1)), the correlation between V_{max}/Amp and \sqrt{k} is very strong ($R=0.98$ for all data), and it is constant across speakers, articulators and gestures (see Table 3 for details). Similarly, in agreement with equation (2), the correlation between $1/T$ and \sqrt{k} is strong too ($R=0.81$ for all data) and it varies just a little across speakers, articulators and gestures (see Table 3 for details). This confirms the fact that LSOMs account well for global dynamical characteristics of speech articulators. However, the damping values calculated from equations (4) and (5) are highly variable. Figure 3 shows the distribution of the normalized damping (damping divided by the critical damping) for data split by speakers and articulators. In all cases the normalized damping is largely distributed between values ranging between zero (undamped system) and 1 (critical damping). In general, the damping variability is huge and cannot be related to any physical properties of the articulatory system. Hence, these damping values should be considered as *ad-hoc* values fitting in both equations (4) and (5) rather than as reliable estimations of actual physical damping characteristics of speech gestures. Given equations (4) and (5) a large variability in damping should be associated with a noticeable variability in stiffness.

To quantitatively estimate this variability, two approaches were used. In the first approach, the standard deviation sd_k and the mean value \bar{k} of the distribution of the stiffness within small intervals (0.5) of variation of V_{max}/Amp was calculated. The relative variability of the stiffness was estimated as the ratio $(2 \cdot sd_k / \bar{k})$. Its variation was studied as a function of V_{max}/Amp ; its average and maximal values were calculated when V_{max}/Amp varied. In a second approach, the same computations were done for 5 ms intervals of the gesture duration T .

Table 3 shows the results. The relative stiffness variability varies little with V_{max}/Amp (the maximal value is close to the average value) and it is located around 20%. This amount of relative variability is found when all data are taken together as well as for data split by speakers, articulators and gestures. The relative stiffness variability is large enough to make any estimation of the stiffness from the ratio V_{max}/Amp quite inaccurate. It is equally large for slow and for fast gestures as illustrated by the top panel of Figure 4 for all data taken together. The relative stiffness variability varies significantly with gesture duration (the maximal value differs strongly from the average one). The average value is located around 60%, with small variations across speakers, articulators and gestures. The bottom panel of Figure 4 shows for all data taken together that the relative variability increases when gesture duration increases. This result was also found when data were split across speakers, articulators or gestures. Hence, stiffness estimations based on gesture durations are even more inaccurate than the estimations based on V_{max}/Amp , and the longer the gesture is, the worse the estimation is.

Discussion

An evaluation was conducted of the LSOM's capacity to account for dynamical properties of speech articulators. The first part of the evaluation was made in the context of an undamped LSOM. It consisted of considering pairs of gestures and comparing the mass-normalized stiffness differences between these gestures as estimated using equations (1) and (2). These equations are two strictly equivalent ways to use the properties of an undamped LSOM to compute mass-normalized stiffness, one based on kinematic properties and the other one based on gesture duration. Our results reveal that both estimations are quite different in a significant number of cases. Cases in which one equation predicts a stiffness increase while the other equation predicts a stiffness decrease are numerous. For long gesture durations (above 200 ms) the percentage of such cases was found to be as high as 50%. Hence, a first important conclusion is that estimations of the mass-normalized stiffness based on an undamped LSOM should be considered with a great deal of caution.

It was also shown that the percentage of opposite estimations is small for short gestures (around 50 ms), but increases when gesture duration increases. A conceivable explanation for the deterioration of the estimation lies in the fact that due to their non-linear elastic properties (see Introduction) tongue and lip stiffnesses intrinsically vary when these soft articulators move. Since long gestures are often associated with larger deformations and displacements, long gestures should undergo greater variation in stiffness over time than short gestures. Consequently, the longer the gesture is, the less correct a modeling based on an undamped LSOM is, which assumes *per se* that stiffness is constant over time.

In the second part of the evaluation, an under-damped LSOM was considered. Mass-normalized stiffness and damping were computed from equations (4) and (5), which link these variables with the ratio V_{max}/Amp and the duration T . In around 20% of the cases, no solution could be found, because the measures of V_{max} , Amp , and T are not compatible with the physical laws of an LSOM. More specifically, the parameter c of equation (6) was found to be systematically smaller than the smallest value ($c=1.57$) possible for an LSOM. Hence, for these data, stiffness estimation from V_{max}/Amp or T has no theoretical justification. Consequently, a prerequisite for the estimation of stiffness from V_{max}/Amp or T is to compute the c value and to remove those data where c is smaller than 1.57.

For the other 80% of the data, the stiffness and damping values computed from equations (4) and (5) depict significant variabilities. These variabilities make stiffness estimations quite inaccurate and, consequently, not very reliable. Just as for the estimation based on an undamped model, the inaccuracy of the estimation based on an under-damped model clearly increases when gesture duration increases. This general result shows that the relation between gesture duration and stiffness becomes always less strong when gesture duration increases. It is consistent with the hypothesis that extrinsic factors contribute to the control of gesture duration, especially when this duration is long and achievable for a broad range of stiffness.

Conclusion

To conclude, this study has shown that in speech production the relation between mass-normalized stiffness, peak velocity, gesture amplitude and gesture duration cannot be reliably described by a unique formulation based on a linear second-order model. This result may originate in the variation in the damping factor across gestures. In addition, other factors such as variations in the stiffness and in the damping factor over time may also contribute to this low reliability. Furthermore, there are some indications that external time specification could significantly contribute to determining gesture durations and the kinematic properties of speech gestures. Hence, inferences about motor control and time control in speech production based on stiffness estimation using a linear second-order model should be considered with caution.

Acknowledgments

This work was funded by a grant from the BMBF and the French-German University to the project PILIOS carried out jointly at Gipsa-lab in Grenoble and ZAS in Berlin. It is dedicated to Dieter Fuchs.

References

- Ackermann, H., Hertrich, I. & Scharf, G. (1995). Kinematic analysis of lower lip movements in ataxic dysarthria. *Journal of Speech and Hearing Research*, *38*, 1252-1259
- Browman, C. P. & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53), New York: Academic Press.
- Browman, C. P. & Goldstein, L. (1990a). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics*, *18*, 411-424.
- Browman, C. P. & Goldstein, L. (1990b). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, *18*, 299-320.

- Gerard, J.M., Ohayon, J., Luboz, V., Perrier, P. & Payan, Y. (2005). Non-linear elastic properties of the lingual and facial tissues assessed by indentation technique: Application to the biomechanics of speech production. *Medical Engineering & Physics*, 27, 884-892.
- Fung Y.-C. (1981). *Biomechanics, mechanical properties of living tissues*. Springer Verlag, New York.
- Gomi, H., Honda, M., Ito, T. & Murano, E.Z. (2002) Compensatory articulation during bilabial fricative production by regulating muscle stiffness. *Journal of Phonetics*, 30, 261-279.
- Jannedy, S., Fuchs, S. & Weirich, M. (2010): Articulation beyond the usual: Evaluating the fastest German speaker under laboratory conditions. In Fuchs, S., Hoole, P., Mooshammer, C. & Zygis, M. (eds.), *Between the regular and the particular in speech and language*, (pp. 205-234). Frankfurt/M, Germany.: Peter Lang Verlag.
- Kelso, J.A.S., Vatikiotis-Bateson, E., Saltzman, E. & Kay, B. (1985). A qualitative dynamical analysis of reiterant speech production: Phase portraits, kinematics and dynamic modeling. *Journal of the Acoustical Society of America*, 77, 266-290.
- Kühnert, B. & Hoole, P. (2004) Speaker specific kinematic properties of alveolar reductions in English and German. *Clinical Linguistics and Phonetics*, 18, 559-575
- McMahon, T.A. (1984). *Muscles, reflexes and locomotion*. New Jersey: Princeton University Press.
- Ostry, D.J. & Munhall, K.G. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America*, 77, 640-648.
- Perkell, J.S., Zandipour, M., Matthies, M. & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America*, 112, 1627-1641.
- Rochet-Capellan, A. & Schwartz, J.-L. (2007). An articulatory basis for the labial-to-coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/. *Journal of the Acoustical Society of America*, 121, 3740-3754.

- Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Exp Brain Res*, 15, 129–144.
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333–382.
- Shiller, D.M., Houle, G. & Ostry, D.J. (2005). Voluntary control of human jaw stiffness. *Journal of Neurophysiology*, 94, 2207-2217
- Van Lieshout, P.H.H.M., Bose, A., Square, P.A. & Steele, C.M. (2007). Speech motor control in fluent and dysfluent speech production of an individual with apraxia of speech and Broca's aphasia. *Clinical Linguistics & Phonetics*, 21, 159-188.
- Xu, Y. & Wang, M. (2009). Organizing syllables into groups – Evidence from F_0 and duration patterns in Mandarin. *Journal of Phonetics*, 37, 502-520.

Figure Captions

Figure 1: Distribution of *OpposEstim* computed for all data as a function of gesture duration T_1 and T_2 .

Figure 2: Distribution of all data in the $(T, V_{max}/Amp)$ plane. Light gray: data for which no solution was found for equations (4) and (5); dark gray: data for which a solution was found for equations (4) and (5). Bold solid line: $V_{max}/Amp=1.57/T$. Dashed-dotted line: $V_{max}/Amp=1.44/T$, accounting for data plotted in light gray. Dotted line: $V_{max}/Amp=1.85/T$, accounting for data plotted in dark gray.

Figure 3: Distribution of the normalized damping (damping divided by the critical damping) solution of equations (4) and (5) as a function of gesture duration for data split by speakers and articulators.

Figure 4: Variation of the relative stiffness variability (see text) as a function of V_{max}/Amp (top panel) and gesture duration (bottom panel) for all data together.

Table 1:

Characteristics of the *OpposEstim* distributions. Speaker “all” corresponds to the analysis of all data together. “Min” and “Max” are the minimal and maximal values of *OpposEstim* in the distribution; “% non_zero” is the proportion of elementary (20 ms x 20 ms) squares of the gesture duration plane (T_1, T_2) in which *opposite estimations* were found; “Slope” is the slope of the linear regression between T_1 and the average value of *OpposEstim* computed around the diagonal in the (T_1, T_2) plane (see text);

Speaker	Max	Min	% non_zero	Slope
all	47.1	2.9	52.2	469.2
sp1	46.7	3.3	46.2	409.5
sp2	36.7	3.3	58.8	781.0
sp3	41.9	3.2	54.8	649.8
sp4	46.7	3.3	58.0	684.5
sp5	41.0	2.56	52.6	809.5
sp6	33.3	2.1	41.5	1031.3
sp7	57.6	3.0	55.9	822.5
sp8	51.5	3.0	56.0	535.7
sp9	60.0	3.3	60.0	1290.5

Table 2:

Global analysis of the results computed from Equations (4) and (5)

Table 2a – Data split by speakers for all gestures together

Table 2b – Data split by articulators and gestures' types for all speakers together

“NoComp”: percentage of data for which no solution is found for Equations (4) and (5)

“c1” c coefficient in $V_{max}/Amp=c/T$ for data compatible with Equations (4) and (5)“c2” c coefficient in $V_{max}/Amp=c/T$ for data not compatible with Equations (4) and (5)

Speaker “all” corresponds to the analysis of all data together

“Articul” refers to the EMMA sensors taken into account

Gest: ‘O’ = opening; C=closing

“dor”: tongue dorsum sensor; ”jaw”: jaw sensor; “llip”: lower lip sensor;

“ttip”: tongue tip sensor

Speaker	%NoComp	c1	c2
all	21.5	1.85	1.44
sp1	22.3	1.88	1.44
sp2	18.4	1.85	1.44
sp3	20.1	1.84	1.44
sp4	13.3	1.89	1.47
sp5	23.1	1.85	1.44
sp6	22	1.80	1.45
sp7	29.5	1.79	1.44
sp8	23.5	1.86	1.42
sp9	18.5	1.89	1.44

2a

Articul	Gest	%NoComp	c1	c2
dor	C	27.4	1.81	1.45
dor	O	29.3	1.81	1.44
jaw	C	30.2	1.76	1.43
jaw	O	23.1	1.80	1.44
llip	C	7.8	1.91	1.48
llip	O	18.8	1.81	1.47
tback	C	35.6	1.81	1.43
tback	O	39.2	1.80	1.42
ttip	C	5.8	1.99	1.46
ttip	O	14	1.89	1.46

2b

Table 3

Relative variations of stiffness as computed from Equations 4 and 5

Table 3a – Data split by speakers for all gestures together

Table 3b – Data split by articulators and gestures' types for all speakers together

“Corre1”: correlation coefficient between \sqrt{k} and I/T “Corre2”: correlation coefficient between \sqrt{k} and V_{\max}/Amp “AvVari1”: average relative stiffness variability when V_{\max}/Amp varies (cf Fig. 4a)“MaxVari1”: the maximum relative stiffness variability when V_{\max}/Amp varies (cf Fig. 4a)

“AvVari2” average relative stiffness variability when T varies (cf Fig. 4a)

“MaxErr2” maximum relative stiffness variability when T varies (cf Fig. 4a)

Speaker “all” corresponds to the analysis of all data together

“Articul” refers to the EMMA sensors taken into account

Gest: ‘O’ = opening; C=closing

“dor”: tongue dorsum sensor; ”jaw”: jaw sensor; “llip”=lower lip sensor;

“ttip” = Tongue tip sensor

Speaker	Cor1	Cor2	AvVari1	MaxVari1	AvVari2	MaxVari2
all	0.81	0.98	21.4	23.9	66.2	118.7
sp1	0.73	0.97	21.9	32.8	67.9	98.7
sp2	0.76	0.97	20.3	25	61.5	119.7
sp3	0.76	0.97	20.9	26.9	65.6	129.2
sp4	0.75	0.97	20.2	24.3	65.6	104.9
sp5	0.82	0.97	22.5	28.3	61.1	132.7
sp6	0.81	0.97	19.7	29.9	54	103.7
sp7	0.85	0.98	19.9	25.1	55.7	86.8
sp8	0.77	0.97	21.3	25.3	61.4	84.6
sp9	0.72	0.97	21.8	26.9	72.4	121.1

3a

Articul	Gest	Cor1	Cor2	AvVari1	MaxVari1	AvVari2	MaxVari2
dor	C	0.81	0.97	20.4	24.9	51.7	72.1
dor	O	0.84	0.98	20.2	26.7	57.9	97.9
jaw	C	0.85	0.98	19.8	27.3	54.4	87.3
jaw	O	0.83	0.98	19.4	22.8	58.3	114.8
llip	C	0.83	0.97	19.8	23.8	52.1	101.6
llip	O	0.81	0.97	19	22.2	51.5	93.4
tback	C	0.87	0.98	20.5	26.4	57.2	102.1
tback	O	0.87	0.98	20.5	26.4	58.6	97.4
ttip	C	0.71	0.97	20.6	24.5	68.1	118.3
ttip	O	0.78	0.97	21	26.7	63.7	116

3b

Appendix: Linear damped second-order model

Mass-normalized stiffness k ; Mass-normalized damping factor b

Position $x(t)$; Velocity $v(t)$; Acceleration $a(t)$

Differential equation of movement:

$$a(t) + b \cdot v(t) + k \cdot x(t) = 0 \quad (\text{i})$$

Initial conditions: $t=0$, $x(t)=A_0$ and $v(t)=0$.

Under these conditions:

$$x(t) = A_0 \cdot \frac{2\sqrt{k}}{\sqrt{4\cdot k - b^2}} \cdot \cos\left(\frac{\sqrt{4\cdot k - b^2}}{2} \cdot t + \varphi\right) \cdot e^{-\frac{b}{2}t} \quad (\text{ii})$$

$$\text{with } \varphi = \tan^{-1}\left(\frac{-b}{\sqrt{4\cdot k - b^2}}\right) \text{ and } \varphi \in \left[-\frac{\pi}{2}, 0\right]$$

$$v(t) = -A_0 \cdot \frac{2\cdot k}{\sqrt{4\cdot k - b^2}} \cdot \sin\left(\frac{\sqrt{4\cdot k - b^2}}{2} \cdot t\right) \cdot e^{-\frac{b}{2}t} \quad (\text{iii})$$

$$a(t) = A_0 \cdot \frac{2\cdot k \cdot \sqrt{k}}{\sqrt{4\cdot k - b^2}} \cdot \cos\left(\frac{\sqrt{4\cdot k - b^2}}{2} \cdot t + \theta\right) \cdot e^{-\frac{b}{2}t} \quad (\text{iv})$$

$$\text{with } \theta = \left(\tan^{-1}\left(\frac{b}{\sqrt{4\cdot k - b^2}}\right) - \pi\right) \text{ and } \theta \in \left[-\pi, -\frac{\pi}{2}\right]$$

The displacement during the first half pseudo-period is assumed to represent the articulator displacement during a gesture. From these equations, the ratio V_{max}/Amp and the movement duration T can be calculated according to the following equations:

$$T = \frac{2\cdot\pi}{\sqrt{4\cdot k - b^2}} \quad (\text{v})$$

$$\frac{V_{max}}{Amp} = \frac{\sqrt{k}}{1 + e^{-\frac{b\cdot T}{2}}} \cdot e^{-b\cdot\frac{T}{2}\left[\frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left(\frac{b\cdot T}{2\cdot\pi}\right)\right]} \quad (\text{vi})$$

Equation (vi) can also be written as follows

$$\frac{V_{max}}{Amp} = \alpha \cdot \sqrt{k} \quad \text{with} \quad \alpha = \frac{1}{1 + e^{-\frac{b\cdot T}{2}}} \cdot e^{-b\cdot\frac{T}{2}\left[\frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left(\frac{b\cdot T}{2\cdot\pi}\right)\right]} \quad (\text{vii})$$

and equation (v) as

$$T = \frac{\beta}{\sqrt{k}} \text{ with } \beta = \frac{\pi}{\sqrt{1 - \frac{b^2}{4 \cdot k}}} \quad (\text{viii})$$

Thus, if b is constant across gestures, $\frac{V_{max}}{Amp} = \alpha \cdot \sqrt{k}$ and $T = \frac{\beta}{\sqrt{k}}$, α and β being constant. If the system is undamped, $b=0$ and then $\alpha=1/2$ and $\beta=\pi$.

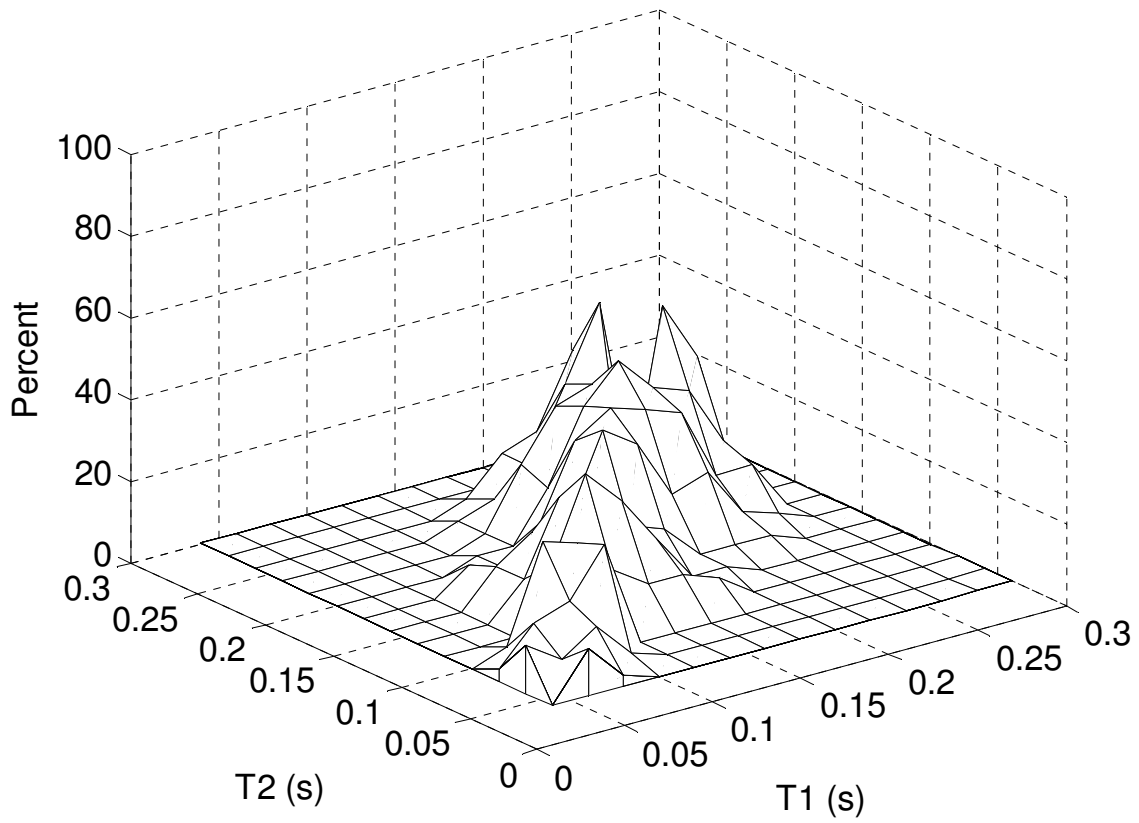


Figure 1

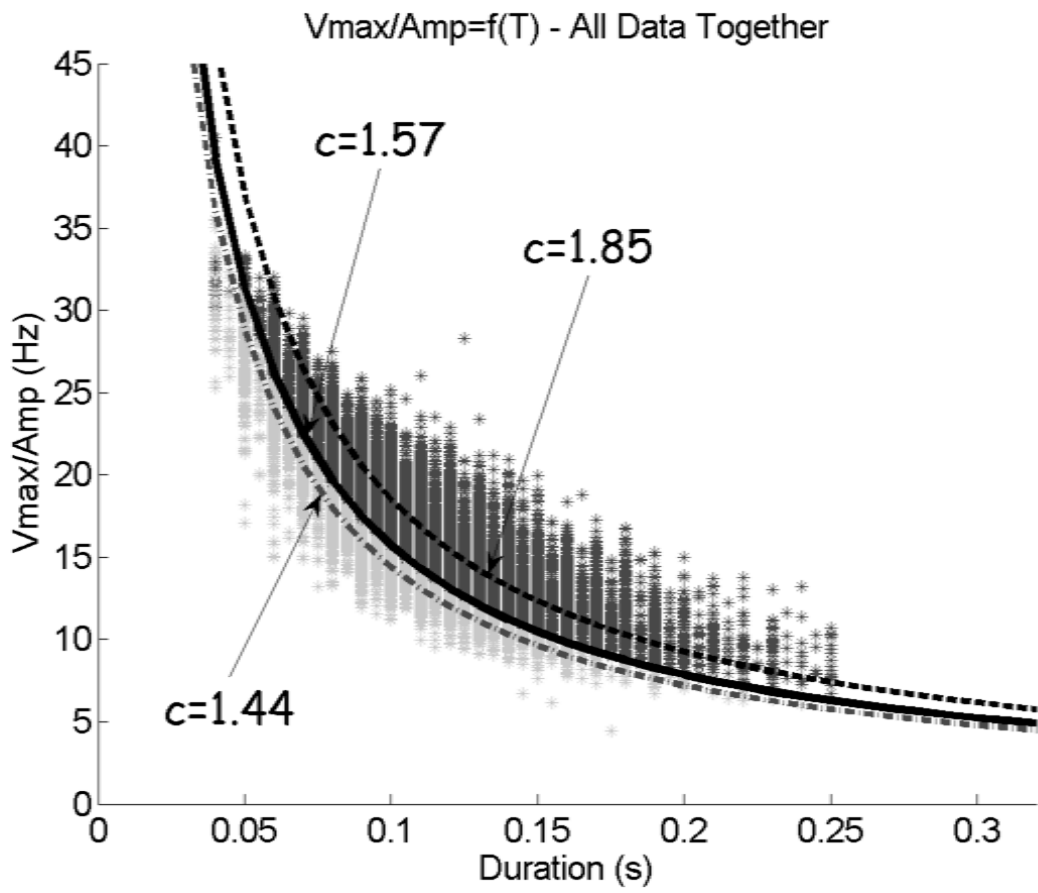


Figure 2

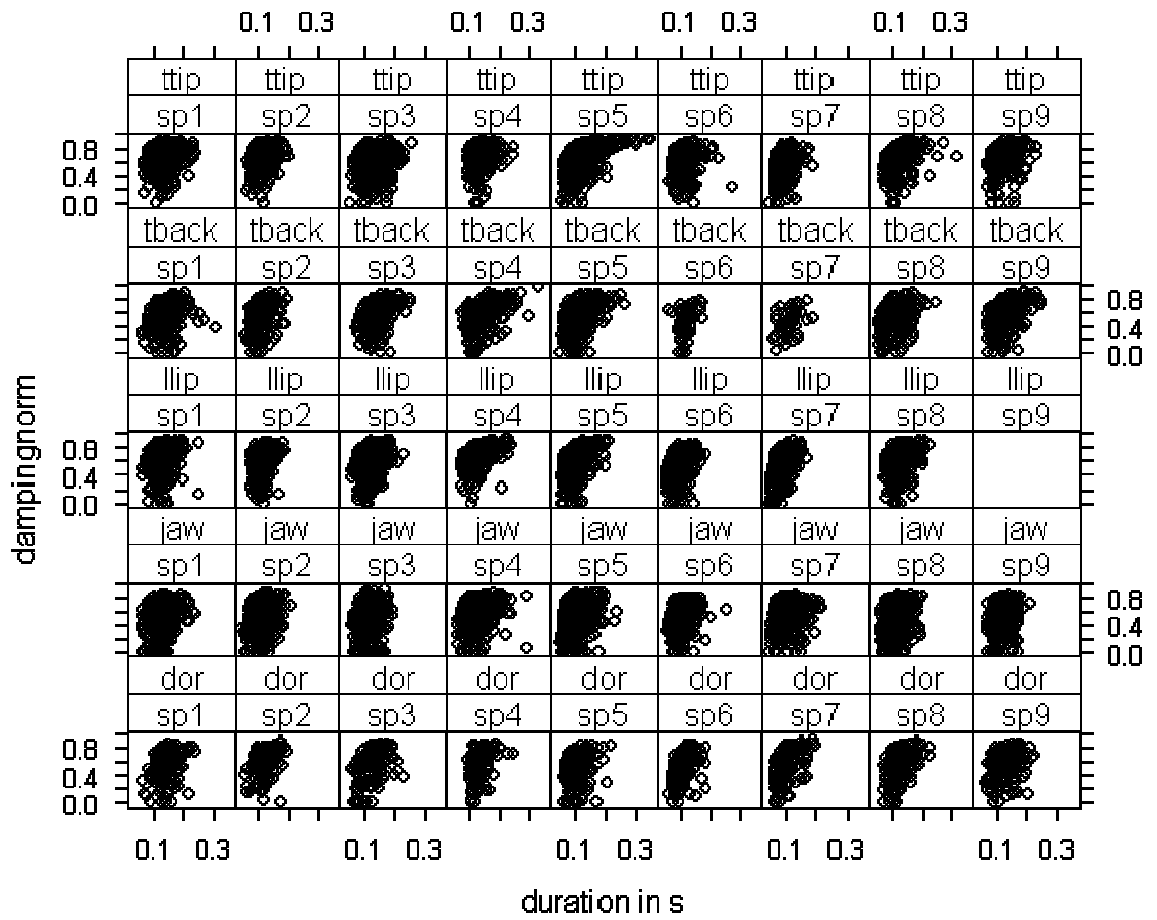


Figure 3

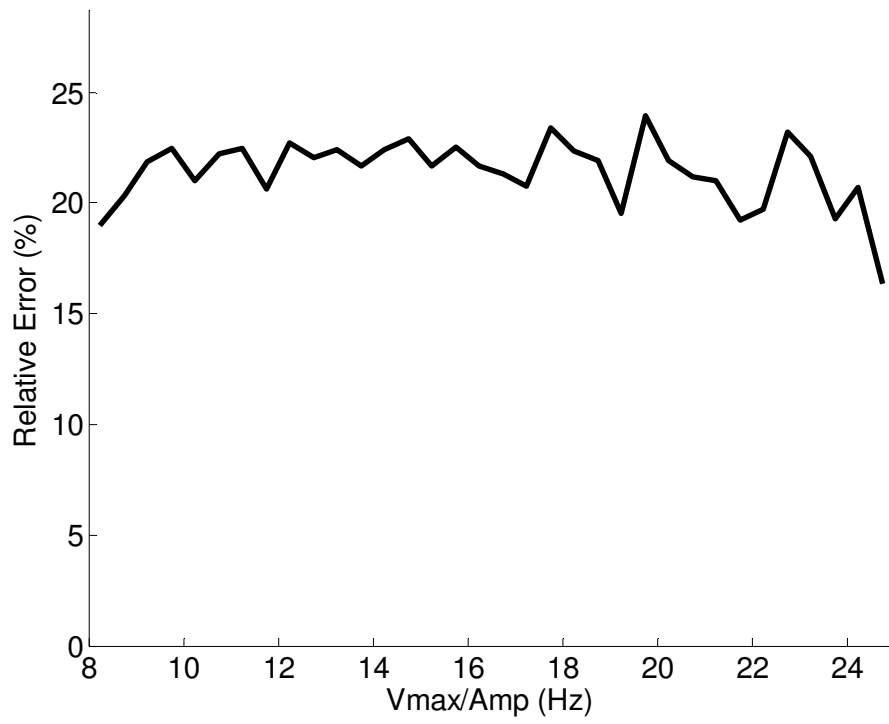


Figure 4.a

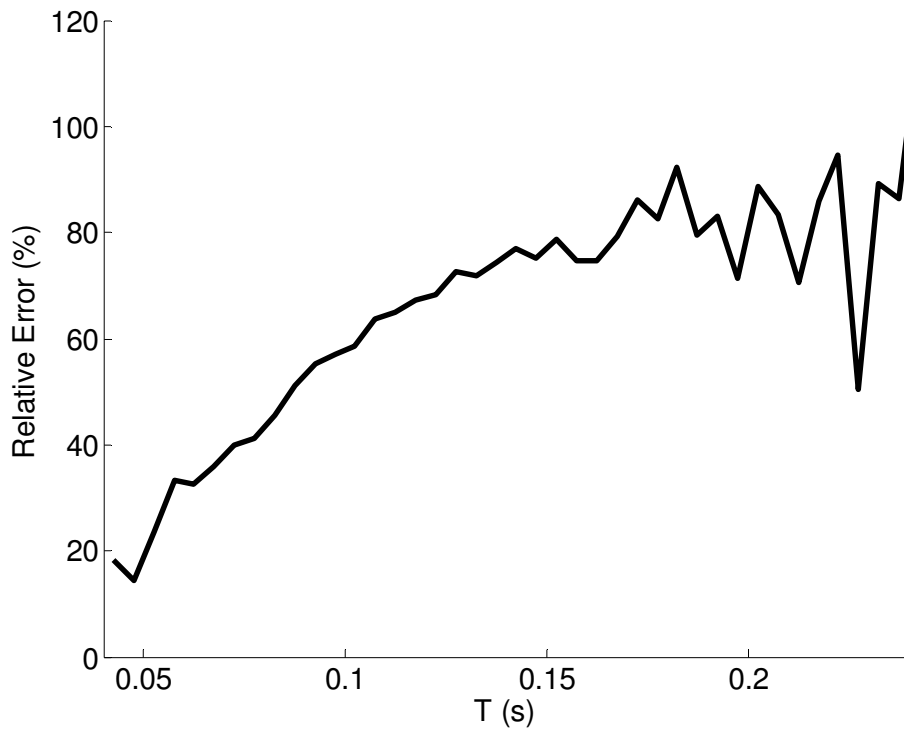


Figure 4.b

hal-00617908, version 1 - 30 Aug 2011