Mathématiques et sciences humaines
Mathematics and social sciences

180 | hiver 2007
Mathématiques et phonologie

A stochastic model for the speech sonority

Un modèle stochastisque en sonorité des langues

Marzio Cassandro, Pierre Collet, Denise Duarte, Antonio Galves et Jesus Garcia

Édition électronique
URL : http://journals.openedition.org/msh/7653
DOI : 10.4000/msh.7653
ISSN : 1950-6821

Éditeur
Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée
Date de publication : 1 décembre 2007
Pagination : 43-55
ISSN : 0987-6936

Référence électronique
Marzio Cassandro, Pierre Collet, Denise Duarte, Antonio Galves et Jesus Garcia, « A stochastic model for the speech sonority », *Mathématiques et sciences humaines* [En ligne], 180 | hiver 2007, mis en ligne le 21 février 2008, consulté le 22 avril 2019. URL : http://journals.openedition.org/msh/7653 ; DOI : 10.4000/msh.7653

Math. & Sci. hum. / Mathematical Social Sciences (45^e année, n° 180, *2007*(4), p. 43–55)

A STOCHASTIC MODEL FOR THE SPEECH SONORITY[1]

Marzio CASSANDRO[2], Pierre COLLET[3], Denise DUARTE[4],
Antonio GALVES[5], Jesus GARCIA[6]

RÉSUMÉ – Un modèle stochatisque en sonorité des langues
Nous étudions des familles de chaînes quantifiées à valeurs réelles. Ces chaînes sont liées par une hypothèse d'existence d'une partition universelle de leur image, telle que la loi de chaque chaîne conditionnée par l'appartenance à un de ses éléments est indépendante de la chaîne. Nous introduisons une nouvelle classe d'estimateurs des points de séparation définissant la partition et démontrons la consistance de ces estimateurs. Nous pouvons alors utiliser ces résultats pour modéliser l'évolution de la sonorité des langues naturelles, sur la base d'un corpus linguistique de 1667 propositions en huit langues différentes. Nous montrons qu'un modèle avec quatre points universaux de séparation décrit bien les données. La notion nouvelle de famille de chaînes quantifiées liées pourrait s'appliquer à d'autres situations dans lesquelles différents agents stochastiques s'expriment par l'intermédiaire du même genre d'interface.

MOTS CLÉS – Chaînes quantifiées, Chaînes stationnaires, Estimation linguistique croisé des points de coupures, Points de coupure universels, Sonorité des langues

SUMMARY – *We study families of bounded real valued tied quantized chains. The chains are tied together by the assumption that there is a universal partition of the range, such that the distribution of the chains, conditioned on each interval of the partition is independent of the chain. We define a new class of cross estimators for the cut-points separating these intervals and prove their asymptotic consistency. We apply our results to model the sonority time evolution of different languages using a linguistic corpus with 1667 sentences from eight different languages. We show that a model with four universal cut-points is in good agreement with the data. The new notion of family of tied quantized chains should be relevant for modeling other situations in which different stochastic agents express themselves using the same type of interface.*

KEYWORDS – Cross-linguistic estimation of the cut-points, Speech sonority, Stationarity chains, Tied quantized chains, Universal cut-points

[1]This work is part of PRONEX/FAPESP's Project *Stochastic behavior, critical phenomena and rhythmic pattern identification in natural languages* (grant number 03/09930-9) and of CNPq's project *Stochastic modeling of speech* (grant number 475177/2004-5).

[2]Dipartimento di Fisica , Università di Roma *La Sapienza*, Piazzale Aldo Moro, 5 , 00185 Roma, Italy, cassandro@roma1.infn.it

[3]Centre de Physique Théorique, CNRS UMR 7644, École Polytechnique, F-91128 Palaiseau Cedex, collet@cpht.polytechnique.fr

[4]Departamento de Estatística, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, CEP 31270-901 Belo Horizonte, MG, Brazil , denisedsma@yahoo.com.br

[5]Instituto de Matemática e Estatística , Universidade de São Paulo, Rua do Matão, 1010, 05508-090 São Paulo SP, Brazil, galves@ime.usp.br

[6]Instituto de Matemática, Estatística e Computação Científica, Unicamp Cidade Universitária *Zeferino Vaz*, 6166 Campinas SP, Brazil, jg@ime.unicamp.br

1. INTRODUCTION

The sonority can be defined as a local index of regularity of the speech signal (see [Galves *et al.*, 2002]). This indexis a function which maps local windows of the acoustic signal on the interval $[0, 1]$. This function assumes values close to 1 in the regions in which the signal presents a regular behavior characteristic of portions of the signal. In contrast, the function will assign values close to 0 to regions characterized by obstruency.

An exploratory analysis of a sample with 1667 sentences from 8 different languages shows that the time evolution of the sonority is quite regular in high level regions and displays strong variations below a certain level. This suggests the modeling of the sonority time evolution of a language by a stochastic quantized chain. Moreover, in this model the quantized chains corresponding to different languages are tied together by the assumption that the distribution of the sonority, conditioned on the fact that it belongs to a given region, is *universal*, i.e. language independent. In particular the partition in regions of sonority is assumed to be language independent. In this model the specific features characterizing each language are expressed by the symbolic chain indicating in which region of sonority the process is at each time step.

This model is linguistically appealing. On one hand the universality of the sonority regions and the corresponding probability distributions mimics the fact that the physiological features of the speech production apparatus are common to all human beings and therefore are language independent. On the other hand the fact that the law of the underlying symbolic chain depends on the language accounts for the specific phonological features characterizing each particular language.

In order to implement the model we need to estimate the universal cut-points separating the sonority regions. In the present paper we introduce a family of cross-linguistic estimators of the cut-points and prove their asymptotic consistency. As far as we know this is a new theoretical result.

Markov quantized chains have been recently considered in the statistical literature (cf. [Bühlmann, 1999], and references therein). However the notion of tied family of quantized chains seems to be new. It is noteworthy that in our model the embedded categorical chain is not assumed to be Markovian as it is usually done in the literature. Actually to prove our theoretical results we only need to assume that the categorical chain is stationary and ergodic.

The paper is organized as follows. In Section 3. we present the linguistic data. In Section 2. we introduce the notion of tied family of quantized chains, define a family of cross-linguistic estimators for the universal cut-points and prove their consistency. In Section 4. we discuss the adequacy of a family of tied quantized chains with four cut-points to model the linguistic data. Final remarks and perspectives are presented in Section 5. The proof of the mathematical results stated in Section 2. is given in the appendix. The data sets and computer codes used in this paper can be obtained from the site [www.ime.usp.br/∼tycho/prosody/sonority/quantized].

2. FAMILIES OF TIED QUANTIZED CHAINS

We will consider a family of stochastic processes $\left\{\left(S_t^l\right)_{t\in\mathbf{Z}} \,:\, l\in\mathcal{L}\right\}$ taking values in the interval $[0,1]$, where $\mathcal{L}$ is a fixed but otherwise arbitrary set. We will assume that these processes are stationary and ergodic. They are tied together by the following assumption.

ASSUMPTION 1. *There exist a positive integer N and an increasing sequence of cut-points $c_0 = 0 < c_1 < \ldots < c_N < c_{N+1} = 1$ and $N+1$ probability measures π_j, $j = 0,\ldots,N$, such that the support of π_j is contained in the interval $I_j = [c_j, c_{j+1}[$ and that at any time step t and for any $l\in\mathcal{L}$ we have*

$$\mathbb{P}\left\{S_t^l \in B | S_t^l \in I_j\right\} = \pi_j(B)\ , \tag{1}$$

where B is any Borel subset of $[0,1]$.

We stress the fact that by assumption, the cut-points c_j and the probabilities π_j, $j = 0,\ldots,N$ are independent of l. In our linguistic application $\mathcal{L}$ will be discrete and will represent a set of natural languages. The intervals I_j will represent regions of different sonority levels.

We introduce the chain $\left(X_t^l\right)_{t\in\mathbf{Z}}$ taking values in the finite alphabet $\mathcal{A} = \{0,\ldots,N\}$ and defined by

$$X_t^l = j \quad \text{if} \quad S_t^l \in I_j\ .$$

The assumptions on $\left(S_t^l\right)$ imply that the chains $\left(X_t^l\right)$ are stationary and ergodic. We introduce the shorthand notation

$$p^l(j) = \mathbb{P}\left\{X_t^l = j\right\}\ .$$

Let $w : [0,1] \to [0,1]$ be a continuous and strictly increasing function with $w(0) = 0$. Given a couple l and l' of different elements of $\mathcal{L}$, we define

$$W^{l,l'}(r) = w(|F^l(r) - F^{l'}(r)|)\ , \tag{2}$$

where $F^l(r) = \mathbb{P}\left\{S_t^l < r\right\}$.

PROPOSITION 1. *Under Assumption 1, assume that each probability π_j has no atom and that its support is the full interval I_j. If $p^l(j) \neq p^{l'}(j)$ for any $j\in\mathcal{A}$, then for any continuous and strictly increasing function w vanishing at the origin, the function $W^{l,l'}(\cdot)$ has a global maximum which is attained at one of the cut-points. In particular, if $N = 1$, then the function $W^{l,l'}$ is unimodal and its maximum is attained at c_1.*

For simplicity, from now on we will assume that this global maximum is unique and will denote by $c^{l,l'}$ the cut-point where the global maximum of the function $W^{l,l'}$ is attained.

PROPOSITION 2. *Under the same assumptions as in Proposition 1, each interval between two zeros of the function $W^{l,l'}$ contains at least one cut-point.*

The reader should note that by definition the function $W^{l,l'}$ has at least zeros at the end points $r = 0$ and $r = 1$.

Propositions 1 and 2 suggest an estimation strategy for the cut-points. First we introduce the empirical counterpart of the function $W^{l,l'}$. For $T \geq 1$, $r \in [0,1]$ and any pair of samples $S_1^l, \ldots, S_T^l$ and $S_1^{l'}, \ldots, S_T^{l'}$, we define

$$\widehat{W}_T^{l,l'}(r) = w(|\widehat{F}_T^l(r) - \widehat{F}_T^{l'}(r)|) \ ,$$

where

$$\widehat{F}_T^l(r) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\{S_t^l \leq r\} \ ,$$

and similarly for $\widehat{F}_T^{l'}$. We define the estimator $\widehat{c}_T^{\,l,l'}$ of $c^{l,l'}$ by

$$\widehat{c}_T^{\,l,l'} = \inf \left\{ v \in [0,1] \,\middle|\, \widehat{W}_T^{l,l'}(v) = \sup_r \widehat{W}_T^{l,l'}(r) \right\} \ .$$

The idea is to define the estimator as the argument of the maximum of $\widehat{W}_T^{l,l'}$. However the fact that the function $\widehat{W}_T^{l,l'}$ is piecewise constant makes it necessary to specify which point is chosen in the interval where the maximum is attained. The following theorem states that $\widehat{c}_T^{\,l,l'}$ provides an asymptotically consistent estimator for $c^{l,l'}$.

THEOREM 1. *Under the same assumptions as in Proposition 1, for any continuous and strictly increasing function w vanishing at the origin, $\widehat{c}_T^{\,l,l'}$ converges almost surely to $c^{l,l'}$, as $T \to +\infty$.*

Theorem 1 allows us to estimate one of the cut-points. When $N \geq 2$, a natural idea would be to repeat the procedure conditioned to each subinterval $[0, c^{l,l'}]$ and $[c^{l,l'}, 1]$. By repeating this procedure iteratively one can hope to identify successive cut-points. A difficulty with a direct application of this idea is that the estimator $\bar{c}$ fluctuates around the true value. Therefore, it is better to consider the maxima of the empirical conditional functions $\widehat{W}_T^{l,l'}(r \,|\, [a_i, b_i])$ where the open intervals (a_i, b_i) form a covering of $(0,1)$.

Formally we define the conditional functions $\widehat{W}_T^{l,l'}(r \,|\, (a_i, b_i))$ as follows. Let

$$\widehat{F}_T^l(r \,|\, (a_i, b_i)) = \frac{\sum_{t=1}^{T} \mathbf{1}\{a_i < S_t^l \leq r\}}{\sum_{t=1}^{T} \mathbf{1}\{S_t^l \in (a_i, b_i)\}} \ .$$

This function is an estimator of

$$F^l(r \,|\, (a_i, b_i)) = \mathbb{P}\left\{ a_i < S_t^l \leq r \,\middle|\, S_t^l \in (a_i, b_i) \right\} \ .$$

One then constructs as above the corresponding empirical conditional function $\widehat{W}_T^{l,l'}(r \,|\, (a_i, b_i))$, and applies Theorem 1 to determine a cut-point inside the interval (a_i, b_i) if any.

The procedure is applied iteratively, refining the intervals (a_i, b_i) until we only get spurious maxima produced by the fluctuations of the empirical distributions. We shall return to the question of how many times one should iterate this procedure in the final section.

3. THE DATA

In Galves *et al.* [2002] an index of local regularity of the speech signal was introduced under the name of *sonority*. This is a mapping of the spectrogram of the acoustic signal into a function of time taking values in the interval $[0, 1]$. At each time step we compute the relative entropy between neighboring normalized columns of the spectrogram. A local average of these relative entropies is then mapped through a fixed decreasing function to define the current value of the sonority.

The definition of the sonority is motivated by the fact that regular patterns characteristic of sonorant spans typically will correspond to sequences of probability measures which are close in the sense of relative entropy. Therefore if the window around time t covers a region of the acoustic signal which is regular, and therefore sonorant, then S_t will be close to 1. In contrast, regions in which the acoustic signal present a chaotic behavior, for instance regions corresponding to stop consonants, will correspond to intervals in which S_t will assume values close to 0, with important variations.

We refer the reader to Galves *et al.* [2002] for a linguistically motivated presentation of the sonority, to Cros *et al.* [2005] for a discussion of the relation between the sonority and the intra-oral pressure and to Cuesta *et al.* [2006] addresses the problem of rhythmic classification of languages using the sonority and the projected Kolmogorov-Smirnov test. As an example, Figure 1 shows the synchronized time evolutions of the pressure (top), of the spectrogram (middle) and of the sonority (bottom) for a piece of a Japanese sentence.

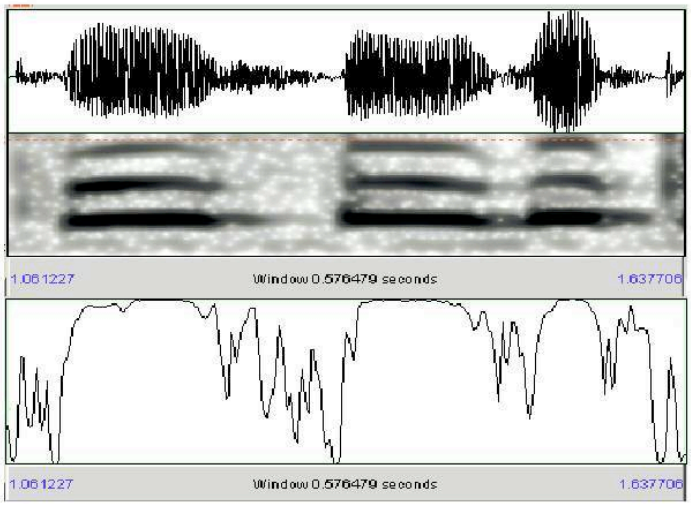

FIGURE 1. Graphs of the acoustic signal (top), spectrogram (middle) and sonority (bottom) for a Japanese utterance. The horizontal axis represents time.

The spectrograms used in the present analysis were produced by the software Praat [www.praat.org]. The computations of the sonority from the spectrogram and some basic statistics were carried out using the Free software Piccolo developed by Jesus Garcia.

The data we use in the present analysis come from two multi-lingual corpora belonging to the *Laboratoire de Sciences Cognitives et Psycholinguistique* (EHESS/ CNRS). The first one was originally recorded by Nazzi *et al.* [1998], with sentences from Dutch, English, French, Italian, Japanese and Spanish. The second one, with sentences in Catalan and Polish, was recorded to be used in the paper by Ramus *et al.* [1999]. The whole corpus consists of sentences recorded by 4 female native speakers of each language, each speaker reading around 50 sentences, controlled with respect to the number of syllables (from 15 to 21), with in total 1667 sentences. The sentences were read in a soundproof booth, were low-pass filtered and digitized at 16 kHz and recorded directly in the hard disk.

An inspection of the corpus shows the same type of behavior for the sonority across languages, namely quite regular time evolutions in the upper sonority region and displaying strong variations below a certain level as exemplified by the time evolution at the bottom of figure 1. It is linguistically appealing to interpret these similarities across languages as an expression of the fact that the physiological features of the speech apparatus are common to all human beings and therefore are language independent. The specific features discriminating different languages should be expressed only in the law of the symbolic chain indicating in which region of the interval $[0, 1]$ the sonority is at each time step.

The above considerations motivate the introduction of the notion of family of tied quantized chains in the next section.

4. A MODEL FOR THE SPEECH SONORITY

Let $\mathcal{L}$ denote the set of 8 languages under consideration. For each $l \in \mathcal{L}$, denote by $\mathcal{U}_l$ the set of recorded sentences from language l in the corpus. It will be convenient to use the representation

$$\mathcal{U}_l = \{(l, i) : i = 1, \ldots, n_l\} \ ,$$

where (l, i) denotes the i^{th} recorded sentences of language l in the corpus and n_l is the total number of recorded sentences of language l.

Denote by $(S_t^{(l,i)})$ the sonority time evolution of sentence (l, i). We assume that the sonority time evolutions corresponding to the different sentences $(l, i) \in \mathcal{U}_l$ are independent realizations of the same stochastic process (S_t^l). We will assume that these processes are stationary and ergodic.

In order to fit a family of tied quantized chains to the linguistic corpus described above, is necessary first to estimate the cut-points using Theorem 1.

Let us start with a descriptive analysis of the set of maxima of the functions $\widehat{W}_T^{l,l'}$, for all 28 possible distinct choices of the languages l and l'. In what follows, in the definition of $\widehat{W}_T^{l,l'}$ we will take $w(x) = x$. To obtain these functions, the empirical distribution $\widehat{F}_T^l$ were calculated using the entire set $\mathcal{U}_l$ of sentences from each language $l \in \mathcal{L}$,, with the formula

$$\widehat{F}_T^l(r) = \frac{\sum_{i=1}^{n_l} \sum_{t=1}^{T_{(l,i)}} \mathbf{1}\left\{S_t^{(l,i)} \leq r\right\}}{\sum_{i=1}^{n_l} T_{(l,i)}} \, , \tag{3}$$

where $T_{(l,i)}$ denotes the length of sentence (l,i) and $T = \sum_{i=1}^{n_l} T_{(l,i)}$. The complete set of 28 graphs can be obtained at the web address [www.ime.usp.br/∼ tycho/prosody/sonority/quantized].

Figures 2, 3, 4 and 5 present the graphs of $\widehat{W}_T^{l,l'}$, with $w(x) = x$, for the pairs of languages (Catalan, English), (Polish, Spanish), (English, Japanese) and (Italian, Japanese), respectively.

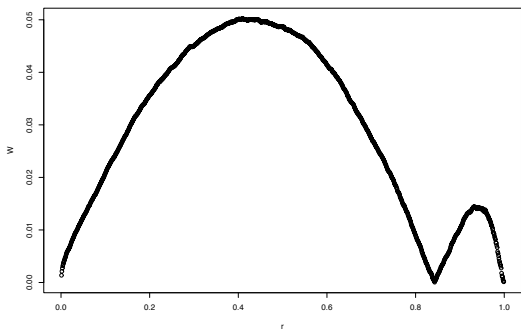

FIGURE 2. Graph of the function $\widehat{W}_T^{l,l'}$ with l = Catalan and l' = English.

Visual inspection of these graphs suggests the existence of at least four cut-points. Indeed, graph 2 shows two maxima attained near 0.44 and near 0.94. The last cut-point reappears in graph 3 which indicates also the existence of a third cut-point near 0.2. Graph 4 suggests the existence of a fourth cut-point in the neighborhood of 0.68. Finally the graph corresponding to the pair (Italian, Japanese) is compatible with the existence of the four cut-points suggested by the previous graphs. However, the scale of this last graph is about ten times smaller than the scale of most graphs, as exemplified by the above mentioned. This could indicate that the probability distributions $p^l(\,\cdot\,)$ and $p^{l'}(\,\cdot\,)$ for l = Italian and l' = Japanese are similar.

Something new appears in graph 6 which suggests the existence of another cut point near 0.05. This seems to be a spurious maximum, produced by an insufficient number of points of small sonority in the sample of French sentences. Indeed the probability of visiting the region of very small sonority is smaller in French sentences than in English sentences (cf. [Galves *et al.* 2002]). The same effect may occur at the other extremity of the interval as exemplified in graph 5, corresponding to the pair (Italian, Japanese). We will return to this point in Section 5.

It turns out that these graphs are representative of the entire set of 28 graphs, in the sense that they show all cut-points, all shapes, and also the few spurious maxima appearing in the other graphs.

To see how the cut-point estimators fluctuate we will use a bootstrap procedure. For each $l \in \mathcal{L}$, let $\xi_i^{l,b}$, $i = 1, \ldots, n$, $b = 1, \ldots, B$ be random variables uniformly distributed in $\{1, \ldots, n_l\}$, where B is a suitable positive integer. Assume that the random variables $\xi_i^{l,b}$ are independent. With these random indexes we will construct

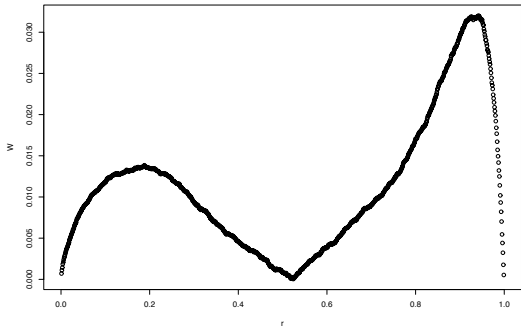

FIGURE 3. Graph of the function $\widehat{W}_T^{l,l'}$ with $l =$ Polish and $l' =$ Spanish.

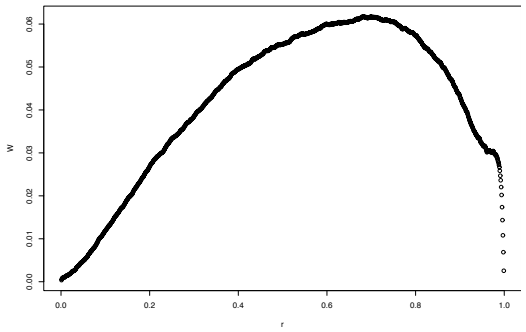

FIGURE 4. Graph of the function $\widehat{W}_T^{l,l'}$ with $l =$ English and $l' =$ Japanese.

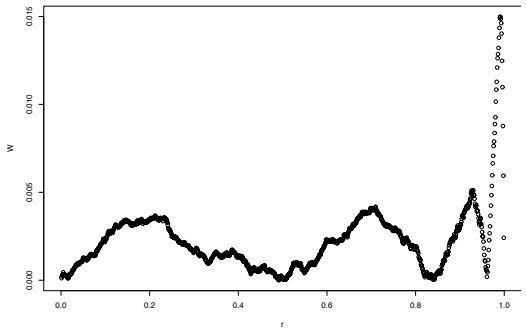

FIGURE 5. Graph of the function $\widehat{W}_T^{l,l'}$ with $l =$ Italian and $l' =$ Japanese.

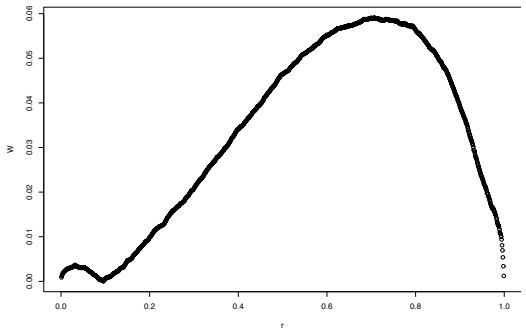

FIGURE 6. Graph of the function $\widehat{W}_T^{l,l'}$ with l = English and l' = French

the bootstrap samples $\{\mathcal{U}_l^{b^*}, l \in \mathcal{L}\}$, $b = 1, \ldots, B$, defined by

$$\mathcal{U}_l^{b^*} = \left\{(l, \xi_i^b) : i = 1, \ldots, n\right\} .$$

We apply to the bootstrap samples the iterative procedure described at the end of Section 2. We used the following covering

$$\left\{(0, 0.3),\ (0.3, 0.55),\ (0.55, 0.8),\ (0.8, 1),\ (0.2, 0.4),\ (0.4, 0.7),\ (0.7, 0.9)\right\} .$$

The cut-points are estimated as follows. In each bootstrap sample we identify the point where the maximum is attained in each interval of the covering (if any). This allows us to identify four clusters of points. The corresponding cut-point is estimated by the median of the cluster

Table 1 summarizes the results obtained with $n = 50$ and for B taking successively the values $100, 200$ and 300. The first column gives the value of B used in the estimation. The second column gives the index of the cut-point considered. The third column shows the estimated values for the cut-points. The fourth and fifth columns give the interquartile distance $(q_3^* - q_1^*)$ and the standard deviation respectively for each cut-point.

We observe that both the interquartile distances and the standard deviations in each cluster are much smaller than the distance between consecutive cut-points. These results are therefore compatible with the existence of four cut-points which moreover are universal since the same cut-points point appear for different pairs of languages.

5. DISCUSSION

In the present paper we introduced the notion of family of tied quantized chains. Even though our original motivation comes from Linguistics, we believe that this could be useful in other fields to model phenomena in which different stochastic agents are constrained to act in the same environment.

The use of tied quantized chains to model the sonority of a set of languages is new. The evidence for our model came from a linguistic corpus which as far as we

B	cut-point index	estimated cut-point	$q_3^* - q_1^*$	standard deviation
100	1	0.187	0.046	0.038
	2	0.448	0.053	0.034
	3	0.672	0.069	0.047
	4	0.930	0.016	0.024
200	1	0.191	0.047	0.040
	2	0.456	0.063	0.038
	3	0.680	0.062	0.045
	4	0.932	0.015	0.025
300	1	0.191	0.049	0.041
	2	0.456	0.050	0.034
	3	0.673	0.067	0.042
	4	0.932	0.014	0.024

Table 1. Estimated cut-points, interquartile distances and standard deviations obtained using bootstrap samples with $B = 100, 200$ and 300.

know has never been entirely submitted to a statistical analysis. Previously, only a small subset with 20 sentences from 8 languages selected from the original Nazzi *et al.* [1998] corpus together with the additional sentences in Catalan and Polish was used in the descriptive analysis performed in Ramus *et al.* [1999]. An inferential analysis of this restricted corpus was given in Duarte *et al.* [2001]. The basis of this analysis was a probabilistic model for the lengths of successive consonantal intervals, represented as independent and identically distributed gamma random variables. This inferential study showed that a model with three different variances for the gamma distributions, one for Dutch, English and Polish, a second one for Catalan, French and Spanish and a third one for Japanese, was compatible with the data as suggested by the original descriptive analysis given in Ramus *et al.* [1999].

The notion of sonority considered here was introduced in Galves *et al.* [2002] as a tool to discriminate between rhythmic classes of languages. The goal was to reproduce in an entirely automatic way, with no need of previous hand labeling, the remarkable empirical results obtained by Ramus, Nespor and Mehler [1999]. While remaining close to the spirit of Ramus *et al.* [1999], this new approach avoids the linguistic difficulties associated to the definition of the statistic parameters considered in Ramus *et al.* [1999] and in Duarte *et al.* [2001]. For a discussion of this issue we refer the interested reader to Galves *et al.* [2002] and to Ramus [2002].

The choice of the value 2.5 for the free parameter appearing in the definition of the sonority was guided by empirical considerations. In fact this is the value which seems to reproduce in a more clear way the three clusters of languages suggested by the empirical analysis presented in Ramus *et al.* [1999].

The idea that the the time evolution of the sonority of different languages is well described by a family of tied quantized chains is linguistically appealing. To fully support this intuition we should go one step further in the statistical analysis and to address the question of the universality of the conditional distribution of the sonority in each interval $\widehat{I}_j = [\hat{c}_j, \hat{c}_{j+1}[$, for $j = 0, \ldots, 4$. This issue seems to require a larger linguistic corpus with longer tokens of speech. Indeed the sonority seems to

have important time correlations at least inside each vocalic and consonantal interval. This could explain the large fluctuations we found in the empirical conditional distributions of the sonority, making it difficult to draw any conclusion concerning this issue with the present corpus.

A natural statistical question arises from our approach, namely the estimation of the number of cut-points. This issue can probably be treated by using a minimum description length principle like BIC (see, for instance [Barron, Rissanen, Yu, [1998]). For isolated quantized chains, a preliminary step in this direction has been suggested in Bühlmann [1999] using the AIC. To implement this approach it is necessary to estimate the likelihood of a sample, and this requires extra conditions on the law of the process. This is outside the scope of the present paper.

The framework considered here is much less restrictive than the one usually found in the recent literature on quantized chains which assumes that the embedded chain is Markov of finite order, together with an independence assumption for the values of the process conditioned on the values of the embedded chain (see [Bühlmann, 1999]). Indeed, to prove Theorem 1, besides Assumption 1, we only assume that for each fixed l, the process $\left(S_t^l\right)$ is stationary and ergodic.

Our model relies on the idea that all the linguistically relevant information is carried by the symbolic chains underlying the sonority time evolution. In particular, the most important linguistic question of the existence of rhythmic classes should be decided using only the properties of the symbolic chains. This issue will be treated in a subsequent paper.

PROOFS

In this appendix we give the proofs of the mathematical results stated in Section 2. The following lemma will appear in all the proofs.

LEMMA 1. *Under the same assumptions as in Proposition 1, the function $F^l - F^{l'}$ is strictly monotone in each interval I_j, $i = 0, \ldots, N$.*

Proof of Lemma. 1 We first observe that for $r \in]c_j, c_{j+1}[$, $j = 0, \ldots, N$, we have

$$F^l(r) = \sum_{k=0}^{j-1} p^l(k) + p^l(j)\ \pi_j([0,r])\ ,$$

where for $j = 0$ the first term is absent, and a similar formula holds for $F^{l'}(r)$. Therefore, for $r \in]c_j, c_{j+1}[$ we obtain

$$F^l(r) - F^{l'}(r) = \sum_{k=0}^{j-1} \left(p^l(k) - p^{l'}(k)\right) + \left(p^l(j) - p^{l'}(j)\right)\ \pi_j([0,r])\ .$$

Since we assumed $p^l(j) \neq p^{l'}(j)$, for any $j \in \mathcal{A}$, we conclude that $F^l(r) - F^{l'}(r)$ is monotone on each interval of the partition of the interval $[0,1]$ with endpoints $0 = c_0 < c_1 < \ldots < c_N < c_{N+1} = 1$. ■

Proof of Proposition 1. The function $F^l - F^{l'}$ is continuous and vanishes at the boundaries $r = 0$ and $r = 1$. Moreover, the hypothesis $p^l(j) \neq p^{l'}(j)$, for any $i \in \mathcal{A}$, implies that $F^l - F^{l'}$ is not identically zero. Therefore it has a maximum and a minimum and at least one of them is not zero.

Lemma 1 implies that any non zero maximum or minimum of the function $F^l - F^{l'}$ is a cut-point. Therefore any maximum of the function $W^{l,l'} = w\big(\big|F^l - F^{l'}\big|\big)$ is also a cut-point. ■

Proof of Proposition 2. If the unique zeros of $\widehat{W}_T^{l,l'}$ are the boundaries $r = 0$ and $r = 1$, then the result follows from Proposition 1. Let us now suppose that there exists $\bar{r} \in (0, 1)$, such that

$$\widehat{W}_T^{l,l'}(\bar{r}) = w\big(\big|F^l(\bar{r}) - F^{l'}(\bar{r})\big|\big) = F^l(\bar{r}) - F^{l'}(\bar{r}) = 0.$$

By Lemma 1 the difference function $F^l - F^{l'}$ is strictly monotone between two consecutive cut-points. Therefore, such interval can contain at most one zero of the difference function. This implies the proposition. ■

Proof of Theorem 1. By hypothesis, the probability measures π_i have no atoms and therefore the functions F^l and $F^{l'}$ are continuous. The compactness of the interval $[0, 1]$ implies that they are actually uniformly continuous on the compact set $[0, 1]$.

The empirical distribution functions $\widehat{F}_T^l$ and $\widehat{F}_T^{l'}$ are by definition non decreasing. By Birkhoff's Ergodic Theorem, for any rational number $r \in [0, 1]$ we have

$$\lim_{T\to+\infty} \widehat{F}_T^l(r) = F^l(r) \text{ and } \lim_{T\to+\infty} \widehat{F}_T^{l'}(r) = F^{l'}(r)$$

almost surely. Therefore by a standard argument, both sequences converge almost surely uniformly in r. Since w is continuous, $\widehat{W}_T^{l,l'}(\cdot)$ converges almost surely uniformly to $W^{l,l'}(\cdot)$, as $T \to +\infty$. The result follows at once. ■

Acknowledgments. We thank Emmanuel Dupoux, Jacques Mehler, Marina Nespor, Janet Pierrehumbert, Frank Ramus and Sharon Pepperkamp for many illuminating discussions. Special thanks to Frank Ramus for making the data used in this paper available to us.

REFERENCES

BARRON A., RISSANEN J. YU B., "The minimum description length principle in coding and modeling. Information theory: 1948-1998", *IEEE Trans. Inform. Theory* 44, 1998, p. 2743-2760.

BÜHLMANN P., "Dynamic adaptive partitioning for nonlinear time serie", *Biometrika* 86, 1999, p. 555-571.

CROS A., DEMOLIN D., FLESIA A.G., GALVES A., "On the relationship between intra-oral pressure and speech sonority", *Interspeech'2005 - Eurospeech*, Lisbon, 2005.

CUESTA-ALBERTOS J.A., FRAIMAN R., GALVES A., GARCÍA J., SVARC M., "Identifying rhythmic classes of languages using their sonority: a Kolmogorov-Smirnov approach" [to appear in *Journal of Applied Statistics*, 2007].

DUARTE D., GARVES A., LOPES N., MARONNA R., "The statistical analysis of acoustic correlates of speech rhythm", *Workshop on rhythmic patterns, parameter setting and language change*, ZiF, University of Bielefeld, 2001.
[http://www.physik.uni-bielefeld.de/complexity/duarte.pdf].

GALVES A., GARCÍA J., DUARTE D., GALVES C., "Sonority as a basis for rhythmic class discrimination", *Speech Prosody 2002*, Aix-en-Provence.
[www.lpl.univ-aix.fr/sp2002/pdf/galves-etal.pdf].

NAZZI T., BERTONCINI J., MEHLER J., "Language discrimination by newborns: towards an understanding of the role of rhythm", *J. Experimental Psychology: human perception and performance* 24, 1998, p. 756-786.

THE PICCOLO PROGRAM, can be downloaded: www.ime.usp.br/~tycho/prosody/piccolo

PRAAT PROGRAM AND MANUALS, can be downloaded: www.praat.org

RAMUS F., "Acoustic correlates of linguistic rhythm: perspectives", *Speech prosody 2002*, Aix-en-Provence, 2002. [www.lpl.univ-aix.fr/sp2002/pdf/ramus.pdf].

RAMUS F., NESPOR M., MEHLER J., "Correlates of linguistic rhythm in the speech signal", *Cognition* 73, 1999, p. 265-292.