Characterisation of rhythmic patterns for text-to-speech synthesis

Plínio BARBOSA and Gérard BAILLY
Institut de la Communication Parlée
U.R.A. CNRS n. 368 INPG/ENSERG - Université Stendhal
46, av. Félix Viallet - 38031 Grenoble CEDEX 1 France

Abstract

This article proposes an alternative rhythmic unit for the syllable: the inter-Perceptual Center group. This group is delimited by events which can be detected using only acoustic correlates [29]. The rhythmic patterns for French are described using this characterisation: we show that realisation of accents is gradual over the trailed accentual group and that this gradual lengthening is needed for perception. A model of repartition of the IPCG duration among its segmental constituents incorporating automatic generation of pauses (emergence and duration) according to speech rate is then described.

1 Introduction

Concatenative speech synthesis offers the most efficient solution to the problem of dealing with large coarticulation effects which are extremely difficult to describe using synthesis-by-rule systems. The search for high intelligibility and naturalness of the resulting synthetic speech leads researchers to use larger and larger synthesis units [34, 14]. This trend is not strongly reflected in the models of prediction of prosody: the synthesis-by-rule approach is widely accepted for generating melodic trajectories and segmental durations. The most common melodic models describe the melodic curve as a discrete sequence of targets connected by interpolating functions [33, 27, 16] which are easily related to a phonological surface structure (see for exemple the HL model [26]). Inductive methods may be used to learn situation- and speaker-specific strategies. Except some attempts cited below, the statistical methods used in this framework take the syllable as the basic unit: f0 "movements" are generated by appropriate phonotactical information given as input, e.g. the position of the syllable in the accentual group [31, 30]. The models of generation of segmental durations are even more local and use generally the phoneme as a basic unit: complex rules eventually trained on large corpus using statistical methods [7, 37] combine co-intrinsic and linguistic factors into additive [24] or multiplicative models [19].

Few studies have tried to generate prosody using global methods. Most of them have been dedicated to the generation of melody using retrieval of melodic "contours" in a hierarchical lexicon indexed by linguistic keys [2] or using dynamical non-linear predictors [21, 35] with powerfull learning abilities.

This study provides evidence of the possible automatic generation of segmental durations using an original rhythmic programming unit, i.e. the inter-PCenter Group. The temporal organisation of these units exhibits monotonous ascending movements with phase resetting which can be easily captured by dynamical non-linear predictors synchronously with f0 movements.

2 About the nature of the rhythmic programming unit

The most widely accepted prosodic unit is the syllable. The quasi-periodic movement of the jaw has earlier given prominence to the role of the syllable in the rhythmic organisation of the speech both in perception and production [13]. As mentioned by D. Hirst [15, p. 34]: "Most of the arguments in favour of the syllable as a unit are really arguments in favour of syllabicity rather than syllabic constituents (...)". In this section we examine an alternative candidate for the rhythmic programming unit.

2.1 The P-Center

The work developed by Marcus [22] about Perceptual-Centers (PC) shows that listeners are capable to adjust consistently the delay between two syllabic occurrences (V versus CV) to perceive such analternance as isochronous. This tendancy to isochrony seems to be one of the components of the rhythmic control system as attested by the discussion about isochrony and isosyllabism [28, 20]. In an experiment described by Pompino-Marschall [29] the PC absolute position is considered to be in the middle of two beat clicks in a sequence syllable/beat or beat/syllable after adjusting the beat sequence to perceive the alternation as isochronous. His results also show that the distance between two consecutive PC determines probably the perception of momentary tempo. Despite the diversity of the consonants the PC location seems to be at the neighbourhood of the vocalic onset.

Recent work [17], however, places the PC at location of significant change of the derivation of the intensity function. Thus PC would migrate left of vocalic onset if a silent preceeds the CV syllable. We follow this finding in our work: the PC location is fixed at the vocalic onset if there is no pause before the syllable and just after the silent period if there is one. The distance between two adjacent PCs has been called the inter PC interval (IPCI) and the segmental units it contains form the inter-PC group (IPCG).

In the following we will demonstrate the optimal consistency of the rhythmic behaviour of the segmental units inside a IPCG.

2.2 The Campbell's model

The elasticity principle [9] in its strongest version says that all segmental durations in a syllable frame are obtained by a same and single factor k by the formula $Dur_i = \exp(\mu_i + k.\sigma_i)$, where μ_i and σ_i are the mean and standard deviation of the log-transformed durations (in milliseconds) of the realisations of the phoneme i. In the second step of generation, a simple algorithm permits to obtain segmental durations by computing the k-factor as $\sum_i Dur_i =$ *syllable duration*. This model of repartition of the syllable duration among its segmental constituents simplifies the automatic generation of segmental duration which then proceeds in two steps [8]: (a) prediction of the syllable duration from phonotactical and phonological information by a statistical model (a multilayer perceptron), (b) application of the repartition model.

It is worth noting that Campbell proposes further modification of the algorithm [8, p. 218]: "This is currently implemented with a use of a decay variable that reduces the effect of any lenghtening on segments earlier in the syllable (...)". This means that the onset receives a weaker lenghtening factor k than the rime. This is confirmed experimentally by our own work [5] showing that the k factors of the onset are positively correlated with the precedent rime whereas it negatively correlates with its own rime.

2.3 Computing the elasticity characteristics for our reference speaker

The elasticity characteristics of the above model μ_i and σ_i are computed using large corpus of natural speech (200 phonetically balanced sentences in [8]). As our model of generation of prosody will be part of a text-to-speech system using concatenative synthesis [4], we used the database of 2012 labelled logatoms designed to gather our "polysounds". The resulting elasticity characteristics for our reference speaker are listed in table 1. Notice that this table captures speaker's articulatory habits such as the unusual lengthening of the realisations of the /s/.

[a]	115(38)	[ɛ]	109(32)	[e]	118(41)	[i]	101(33)	[χ]	107(36)	[x]	116(29)
[y]	105(30)	[ɔ]	110(32)	[o]	125(34)	[u]	108(32)	[j]	94(27)	[w]	95(25)
[h]	92(20)	[ã]	136(42)	[ɔ̃]	138(42)	[ɛ̃]	134(39)				
[p]	125(38)	[t]	128(39)	[k]	114(35)	[b]	92(22)	[d]	94(25)	[g]	87(23)
[f]	143(32)	[s]	163(36)	[ʃ]	144(26)	[v]	106(30)	[z]	116(25)	[ʒ]	104(26)
[m]	104(24)	[n]	102(24)	[g]	110(26)	[r]	94(30)	[l]	93(27)		

Table 1: Means (and standard deviations) of the durations (in ms) of the realisations of the French phonemes for our speaker.

3 Characterisation of rhythmic patterns

3.1 The corpus

We recorded a corpus of 88 sentences pronounced at two speech rates (slow and normal [1]) by our reference speaker. Segmental durations were determined semi-automatically using the Temporal Decomposition algorithm described in [4]: first, centres of phoneme realisations are marked by hand and then, segmental boundaries are set automatically. In order to study the relation between rhythmic and linguistic structure, this corpus was labelled using a simple hierarchical model of prosodic marking [3].

3.2 The rhythmic patterns

The sequences of k-factors computed for each IPCG can be easily segmented into prosodic contours characterized by:

(a) monotonously increasing durational contours marking prosodic boundaries: the end of the contours coincide always with a prosodic marker and the amount of target lengthening correlates with the supposed strength of the marker (see Table 2). Please note that

[1]The speaker was initially instructed to utter at three different rates i.e. slow, normal and fast. But no significant variation in tempo was found between normal and fast! This justifies a posteriori the use of synthetic stimuli in Section 5.

the hierarchy of markers induced by mean target lengthening also holds for the pause insertion rate;

(b) resetting near to zero value after accent realisation;

(c) these contours often exhibit an likely exponential increase: these rhythmic "movements" are generated [2] with a positive acceleration towards the target lengthening.

	Marker				
	ID	IT	RD	SD	LD
k-values	4.4 (0.9)	3.6 (1.1)	1.7 (0.9)	1.3 (0.9)	0.4(0.4)
pauses (%)	96 (26/27)	92 (106/115)	28 (23/81)	19 (17/91)	-

Table 2: Mean values (and standard deviations) of k-factors associated with prosodic boundaries (for meaning of markers see [3]) according to the prosodic marker for the slow rate. The percentage of boundaries followed by a pause is also given (with the number of actual occurrence).

The properties mentioned above are specific to the IPCG and clearly not for the syllable (Fig. 1). Two questions were then asked: (1) are continuously increasing patterns needed to the perception of accentuation or they are just an artefact of production constraints? (2) Is this typical configuration needed to the perception of any kind of isochrony in French connected speech? A perception experiment was elaborated to answer these questions.

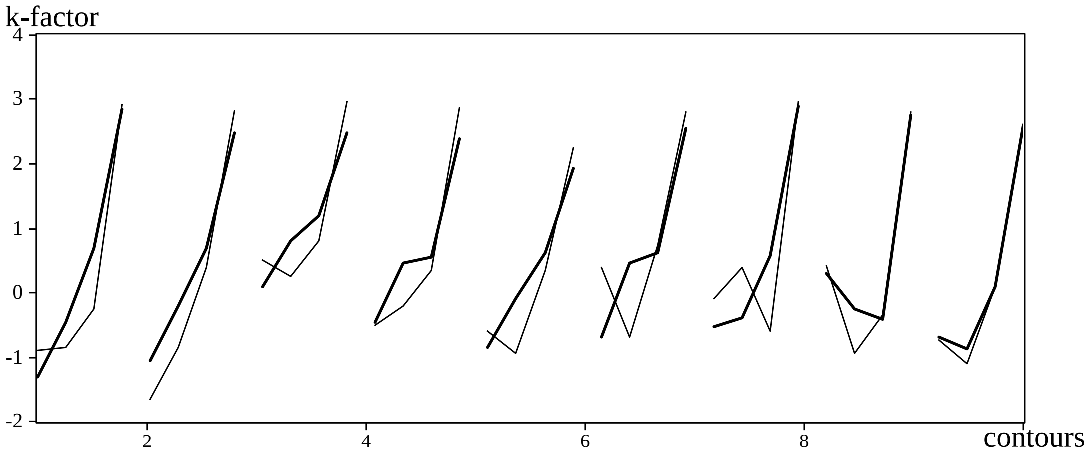

Figure 1: Superposition of k-factors for the syllable and the IPCG (thick lines). These selected contours have 4 syllables and all end with the ID mark.

[2]The tendency of the gradual lengthening over the accentual group in French has been already mentioned in [25]

4 The perceptual relevance of rhythmic patterns

4.1 Method

A preference test was designed to show the perceptual relevance of this production phenomenon [38]. Ten pairs of sentences were generated using a high-quality analysis-resynthesis system [10, 4]: the A configuration was obtained by calculating the successive k-factors of all IPCGs in a sentence. In the B configuration only the k-factors ending an accentual group (cued by a marker) are kept to their original values whereas all others were set to 0. Segmental durations in the two configurations A and B were then computed by applying the repartition algorithm. Please note that (a) segmental durations in A are different from the ones in the natural sentences since k is averaged in each IPCG and (b) on the other hand, the VO timing is identical in both ones since IPCG durations are the same. Silent pause durations and all other parameters were unchanged. An example of computed k-factors for the sequence of IPCGs (A configuration) and the the sequence of k-factors to be tested (B configuration) of an utterance is showed in Fig. 2.

Eleven subjects working in the laboratory but not in synthesis domain participated in this perception experiment. Each session lasted between 10 and 15 minutes. In a test ABBA the subjects were asked to select the most natural utterance. Listening may be repeated twice. At the end of the two possible repetitions they could use a jocker to indicate that they were not able to decide.

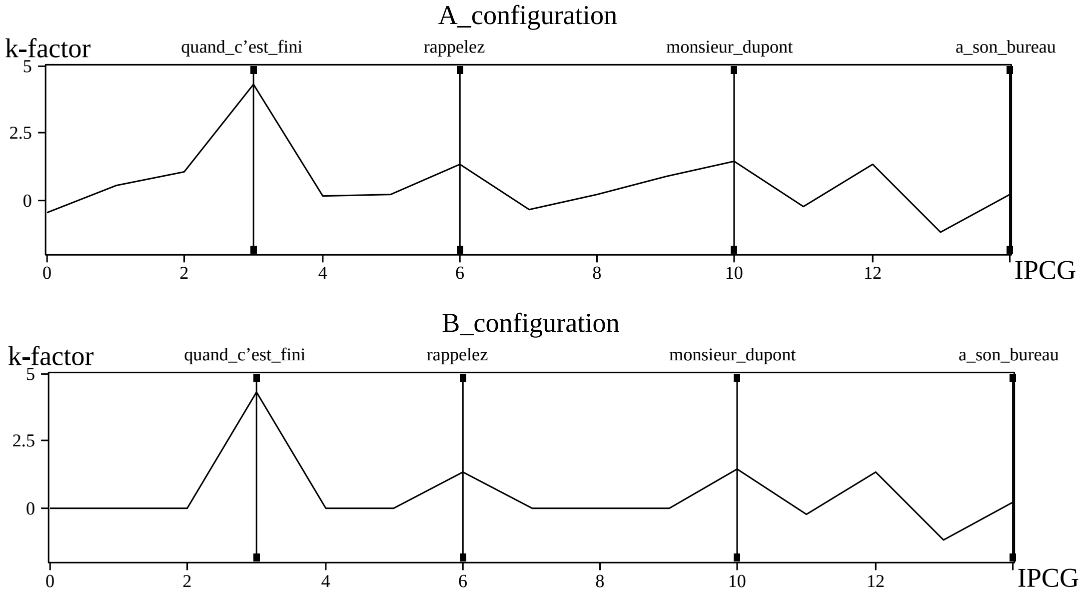

Figure 2: Example of A and B configurations for the sentence: "Quand c'est fini, rappelez Monsieur Dupont à son bureau.". Discrete values of k are connected by lines for sake of visibility. Prosodic groups are cued by vertical bars. The k of the last IPCG of the utterance is taken equal to the k-factor of the vowel.

4.2 Results

All subjects considered, scores are 65% of A configuration, 20% of B configuration and 15% of doubt in choosing one of the utterances. We have then 77% of A configurations over A or

B choices. Individual scores can be seen in Fig. 3.

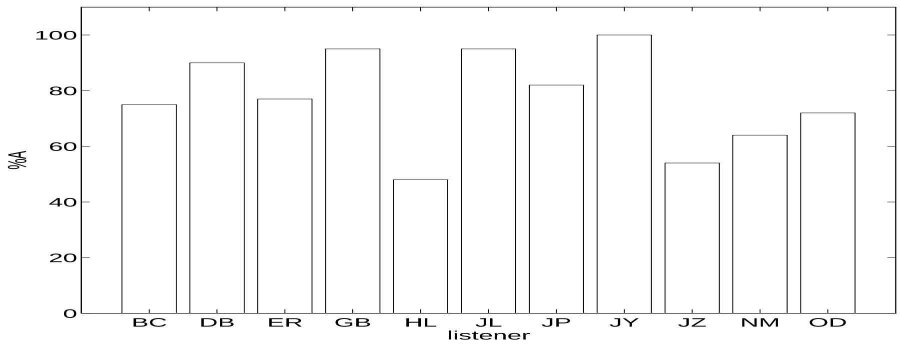

Figure 3: Mean scores (across presentation order) of A selection by subject.

There were no significant effect of presentation order. All listeners agree on the difficulty of the task: "During the experiment I thought I have changed the criterion of utterance choice. In the beginning, I have chosen the ones were more constant concerning the rhythm" (JL), "I have chosen utterances that were the most constant at rhythm level." (OD), "Are you sure they are not identical?" (JY) !!

4.3 Discussion

The clear general preference for the gradual rhythmic pattern lead us to think that this configuration is necessary to the accent perception: lack of lengthening of previous IPCGs sounds abrupt. The *internal clock* (see below) hypothesis may explain the subjects's perceptual behaviour : shorter IPCIs are cues of an unexpected local acceleration. Gradual lengthening contributes to the perception of isosyllabicity [11, 20]. But there is no implicit conclusion that human beings use k-factors to control rhythmic patterns.

5 Incorporating the automatic generation of pauses

5.1 The corpus

A five-rate, 20-sentence corpus was recorded in order to evidence the influence of pause insertion on the rhythmic structure. In order to study the PC phenomenon all sentences of the corpus are CV sequences (all pauses were preceded by a vowel). The analysis of this second corpus confirms for all rates the general trend of rhythmic patterns: 80% of them are monotonously increasing sequences of k-factors.

This experiment aims at developing an automatic generation of duration including the pause phenomenon. To avoid overlaping of speech rates, as noted in section 3.1 the speaker was asked to answer to interrogative synthetic utterances with predefined sentences. These synthetic utterances were obtained roughly using our text-to-speech system by multiplying μ_i and σ_i by a phonation factor [39].

5.2 Characterising speech rate by an internal clock

Our basic assumption of how PC beats are generated and perceived is that the IPCIs are obtained by a deformation from an *internal clock* [1]. This biological clock is the reference for isochrony and the release mechanism of the motor actions [36]. Results of studies on coordination between repetitive tapping and periodic external stimuli tend to reforce this assumption [32]. This internal clock would be a rhythmic attractor : the accentual movement would consist in indicating that "something has happened" as pointed out by the phase difference between the two clocks. Pause phenomenon will the ease the return in-phase of these two clocks [12].

5.3 Analysing pause insertion

Active control of the clock duration is an attractive way for characterising speech rate. But the choice of an internal clock duration is particularly delicate. How to choose an unaccented unit if lengthening is gradual over the prosodic group? Taking into account that: (a) the last IPCG in the prosodic group is clearly the main lengthened unit; (b) if there is a pause, we cannot separate silent and sound intervals : they are elements of the same phenomenon [11]; (c) the first IPCG of a prosodic group is often shortened, the internal clock durations are computed for each utterance as the mean among the nonterminal (excluded prepausal) IPCGs.

This characterisation of speech rate aims to analyse the emergence and durations of actual pauses according to prosodic structure and speech rate. Although most phonological models would assume that the prosodic structure remains the same despite rate changes [23] there is no clear proposals for modelling pause emergence. To our knowledge all text-to-speech systems generate pauses at a phonological level according to ad-hoc phonological and phonotactical constraints. We think that pause emergence has to be solved at the performance level and thus is partly speaker-specific as the parameters of the elasticity model.

We studied the relation between k-factors of vowels computed by including silent in the duration of the given vowel (k_p) and the corresponding k-factors computed only for the vowel duration (k_s). A quasi-asymptotic curve was obtained (see Fig. 4). Regression line of the log-transformed k-factors (previously shifted by the minimum value of k) has given: $(k_s + 5) = (k_p + 5)^{0.59} \exp(0.72)$ for $k_s \neq k_p$. The critical k (0.83) is the point where the regression line crosses the straight line $k_s = k_p$. Please note that a critical k depending on the speech rate do not significantly differ from the average given above nor improves the prediction of location and duration of the pauses.

5.4 Generating sound and silent durations

A minimum silent duration was set for each speech rate as a result of analysis on the corpus.

The repartition algorithm proceeds in two steps: (a) computation of the k-factor for a given IPCG (k_g) and obtention of the corresponding k_s by the regression formula above (by setting $k_p = k_g$), if k_g is greater than the critical k. We then calculate the segmental durations and add them up: the difference to the IPCG duration is assigned to the silent; (b) if the silent duration is greater than the minimum the procedure is over, if not, no duration is assigned to the silent and the k-factor to compute the corresponding segmental durations is k_g. If k_g is lesser than the critical k, the segmental durations are computed using k_g as the k-factor and no duration is assigned to the silent.

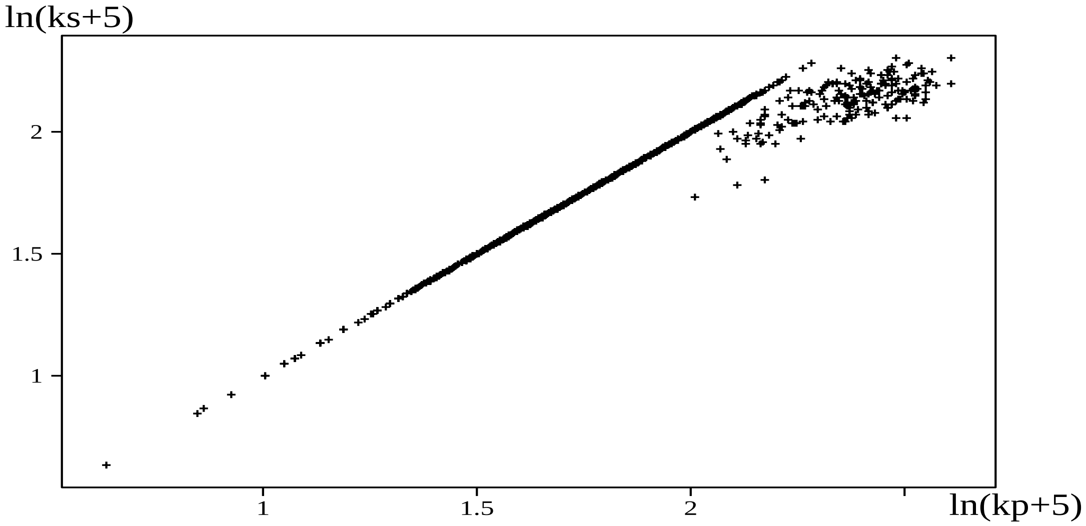

Figure 4: Scatterplot of log-transformed k-factors of vowel versus vowel+silent for all speech rates.

Results of segmental duration generation were computed for three different rhythmic units (see Table 3). Prediction errors are similar between the rhythmic units for vowels but for the consonants and silent durations the PC unit seems to give the best results: lesser mean and ratio standard deviation/mean values than the other two units (see differences in fast and very fast speech rates for example).

rate	clk.	*Vowels*			*Consonants*			min.	*Pauses*		
	(ms)	SY	VO	PC	SY	VO	PC	(ms)	SY	VO	PC
v. slow	360	34(32)	35(32)	34(29)	43(46)	42(43)	37(46)	75	67(58)	76(62)	55(53)
slow	325	38(40)	42(41)	36(32)	32(32)	34(31)	28(26)	67	73(52)	73(52)	73(56)
normal	270	26(26)	30(38)	29(23)	27(26)	28(26)	25(22)	57	54(37)	64(36)	56(35)
fast	210	17(19)	18(19)	19(20)	21(25)	21(27)	19(21)	51	75(77)	68(99)	49(78)
v. fast	165	12(12)	11(13)	12(14)	13(13)	12(17)	11(12)	59	59(33)	93(28)	44(20)

Table 3: Means (and standard deviations) of prediction errors of three sound classes for different rhythmic units using the modified Campbell's model. Average clock durations show that rates are effectively different. The units are: the syllable (SY), the inter-Vocalic Onset group (VO) and the inter-Perceptual Center group (PC). The thresholds are used in the determination of silent pauses.

5.5 Evaluation

It is important to note that no constraints were imposed on the location for the silents, i.e. all syllables were potentially candidates for receiving a pause. Pauses thus "emerge" as a result of the repartition algorithm described above.

Only few silent durations were assigned to positions that were not associated to silent in corresponding natural utterance (see Table 4) but all silent positions were assigned to a latent location for accent realisation (a prosodic marker). The only silent inside a word (in the slow rate) was after the syllable [sɔ] in [sɔmɛ] ("sommet"). After have verified that the [m] duration was very long (291 ms) we realized that the speaker had in fact pronounced a geminated [m] (that is possible in French). We then resegmented "sommet" as [sɔmmɛ] and the algorithm did not place a silent in that position anymore.

rate	clock (ms)	loc. errors	actual silents
v. slow	360	15	99
slow	325	13	64
normal	270	11	32
fast	210	-	26
v. fast	165	1	5

Table 4: Number of errors obtained by the repartition algorithm in placing the silents. The algorythm never associates a nonsilent to a silent in the natural utterance: it generates around 15 % of additional pauses but always located at prosodic boundaries. The number of actual silents by speech rate in the natural utterances is also given. There are 285 IPCGs by rate.

6 Perspectives

As in Campbell's model, our segmental duration predictor [6], proceeds in two stages. Significant differences have to be noticed: (a) a sequential network ([18]) constrained by an internal clock generates timing of PC locations; (b) then the IPCIs are shared between the IPCG constituents according to the repartition model which includes now the emergence of pauses. If the em internal clock is effectively a rhythmical attractor phase resetting has to occur frequently. Our first measurements do not show clear quantisation of the durations of accentual groups (phonation+eventual following silent pause) by the clock as evidenced by [12]. We are currently working on a more robust estimation of the clock duration.

The integration of generation of melody in our predictor is being henceforth planned since both melody and duration will be described by global patterns in our framework. An alternative to this *control* approach could be to use a lexicon of multiparametric contours as proposed by Aubergé [2]. Although the latter approach is more compatible with the concept of prosodic prototypes accessed by linguistic and paralinguistic keys, the control approach has be proven to be more efficient [35].

Acknowlegments

The first author was supported by CNPq-Brasilia/Brasil. We thank especially Frédéric Bimbot which offered his voice and our listeners which offered their ears to such a difficult perception task.

References

[1] Allen, G. Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3:75–86, 1975.

[2] Aubergé, V. Developing a structured lexicon for synthesis of prosody. In Bailly, G. and Benoît, C., editors, *Talking Machines: Theories, Models and Designs*, pages 307–321. Elsevier B.V., 1992.

[3] Bailly, G. Integration of rhythmic and syntactic constraints in a model of generation of French prosody. *Speech Communication*, 8:137–146, 1989.

[4] Bailly, G., Barbe, T., and Wang, H. Automatic labelling of large prosodic databases: tools, methodology and links with a text-to-speech system. In Bailly, G. and Benoît, C., editors, *Talking Machines: Theories, Models and Designs*, pages 323–333. Elsevier B.V., 1992.

[5] Barbosa, P. and Bailly, G. Generating segmental duration by p-centers. In Auxiette, C., Drake, C., and Gérard, C., editors, *Fourth Rhythm Workshop: Rhythm Perception and Production*, pages 163–168. Ville de Bourges, Bourges - France, 1992.

[6] Barbosa, P. and Bailly, G. Génération automatique des p-centers. In *XIXe JEP - Bruxelles*, pages 357–361, 1992.

[7] Bartkova, K. and Sorin, C. A model of segmental duration for speech synthesis in french. *Speech Communication*, 6:245–260, 1987.

[8] Campbell, W. Syllable-based segmental duration. In Bailly, G. and Benoît, C., editors, *Talking Machines: Theories, Models and Designs*, pages 211–224. Elsevier B.V., 1992.

[9] Campbell, W. and Isard, S.D. Segment durations in a syllable frame. *Journal of Phonetics*, 19:37–47, 1991.

[10] Charpentier, F. and Moulines, E. Pitch-synchronous waveform processing techniques for text-to-speech using diphones. *Speech Communication*, 9(5-6):453–467, 1990.

[11] Duez, D. *Contribution à l'étude de la structuration temporelle de la parole en français.* PhD thesis, Université de Provence, 1987.

[12] Fant, G. and Kruckenberg, A. Preliminaries to the study of swedish prose reading and reading style. STL-QPSR 2, 1–80, 1989.

[13] Fraisse, P. *La psychologie du rythme.* Presses Universitaires de France, Paris, 1974.

[14] Hauptmann, A.G. Speakez: a first experiment in concatenation synthesis from a large corpus. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 3, pages 1701–1704, Berlin, 1993.

[15] Hirst, D. Peak, boundary and cohesion characteristics of prosodic grouping. *Working Papers of Lund University*, 41:32–37, 1993.

[16] Hirst, D.J. and di Cristo, A. *Intonation systems: a survey of twenty languages.* Cambridge University Press, Cambridge, forthcoming.

[17] Howell, P. Prediction of p-center location from the distribution of energy in the amplitude envelope. *Perception and Psychophysics*, 43:90–93, 1988.

[18] Jordan, M.I. Motor learning and the degrees of freedom problem. In Jeannerod, M., editor, *Attention and Performance*, volume XIII. Lawrence Erlbaum, Hillsdale, NJ, 1990.

[19] Klatt, D.H. Linguistic uses of segmental duration in english : acoustic and perception evidence. *Journal of the Acoustical Society of America*, 59:1208–1221, 1976.

[20] Lehiste, I. Isochrony reconsidered. *Journal of Phonetics*, 5:253–263, 1977.

[21] Ljolje, A. and Fallside, F. Synthesis of natural sounding pitch contours in isolated utterances using hidden markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34:1074–1080, 1986.

[22] Marcus, S.M. Acoustic determinants of Perceptual center (p-center) location. *Perception and Psychophysics*, 30(3):247–256, 1981.

[23] Monnin, P. and Grosjean, F. Les structures de performance en français : caractérisation et prédiction. *L'Année Psychologique*, 93:9–30, 1993.

[24] O'Shaughnessy, D. A multispeaker analysis of durations in read French paragraphs. *Journal of the Acoustical Society of America*, 76(6):1664–1672, 1984.

[25] Pasdeloup, V. Durée inter-syllabique dans le groupe accentuel en français. In *XIXe Journées d'Etudes sur la Parole*, pages 531–536, 1992.

[26] Pierrehumbert, J. *The phonetics and phonology of English intonation.* PhD thesis, MIT, 1980.

[27] Pierrehumbert, J. Synthetizing intonation. *Journal of the Acoustical Society of America*, 70(4):985–995, 1981.

[28] Pike, K. *The intonation of American English.* University of Michigan Press, Ann Arbor, 1945.

[29] Pompino-Marschall, B. On the psychoacoustic nature of the p-center phenomenon. *Journal of Phonetics*, 17:175–192, 1989.

[30] Sagisaka, Y. On the prediction of global fo shapes for japanese text-to-speech. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1:325–328, 1990.

[31] Scordilis, M. and Gowdy, J. Neural network based generation of fundamental frequency contours. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 219–222, 1989.

[32] Semjen, A., Schulze, H.H., and Vorberg, D. Temporal control in the coordination between repetitive tapping and periodic external stimuli. In Auxiette, C., Drake, C., and Gérard, C., editors, *Fourth Rhythm Workshop : Rhythm Perception and Production*, pages 73–78, Bourges, 1992.

[33] t' Hart, J. and Collier, R. Intonation by rule: a perceptual quest. *Journal of Phonetics*, 1:309–327, 1973.

[34] Takeda, K., Abe, K., and Sagisaka, Y. On the basic scheme and algorithms in non-uniform unit speech synthesis. In Bailly, G. and Benoît, C., editors, *Talking Machines: Theories, Models and Designs*, pages 93–105. Elsevier B.V., 1992.

[35] Traber, C. Fo generation with a database of natural fo patterns and with a neural network. In Bailly, G. and Benoît, C., editors, *Talking Machines: Theories, Models and Designs*, pages 287–304. Elsevier B.V., 1992.

[36] Turvey, M., Schmidt, R., and Rosenblum, L. Clock and motor components in absolute coordination of rhythmic movements. *Haskins Laboratories Status Report on Speech Research*, pages 231–242, 1990.

[37] van Santen, J. and Olive, J. The analysis of contextual effects on segmental durations. *Computer, Speech and Language*, 4:359–390, 1990.

[38] Viviani, P. and Stucchi, N. Motor-perceptual interactions. In Requin, J. and Stelmach, G., editors, *Tutorials in Motor Behavior II.* Elsevier, The Netherlands, 1991.

[39] Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. Segmental durations in the vicinity of prosodic boundaries. *Journal of the Acoustical Society of America*, 91(3):1707–1717, 1992.