Zellner Keller, B., & Keller, E. (forthcoming). The chaotic nature of speech rhythm. In Ph. Delcloque and V. M. Holland (Eds). Integrating Speech Technology in Language Learning. Swets & Zeitlinger

The Chaotic Nature of Speech Rhythm: Hints for Fluency in the Language Acquisition Process

B. Zellner Keller, & E. Keller.
Brigitte.ZellnerKeller@imm.unil.ch

Abstract

The acquisition of speech rhythm and speech fluency are important components of a language acquisition process. This article reviews issues of fluency and speech rhythm, based on empirical, experimental and mathematical evidence. An integrated multilevel model of temporal control for speech is motivated. Additional insights are obtained through a comparison with chaotic systems.

Introduction

An important aspect of any language acquisition process concerns speech *fluency*. Speech fluency designates a dimension that qualifies speakers with respect to their ease of delivering speech (Pfauwadel, 1986; Zellner, 1994). It will be argued in this article that the capacity to speak spontaneously and effortlessly, with smooth transitions and at a relatively fast speech rate, results from the adequate coordination of various processing operations, since speaking is a multi-level cognitive process based on the synchronised activation of various cortical areas (*e.g.*, motor area, perception areas, language areas) operating at different speeds (*e.g.*, Kelso, *et al.*, 1986; Keller, 1990). This coordination — or fluency — is assumed to be directly related to the *temporal dimension of speech* (Zellner, 1998). It means that verbal ease is assumed to occur under conditions of temporal coherence, that is, when speech events follow each other without collision. To our mind, the didactics of language learning should take this aspect into account.

To better understand the temporal organisation of speech as well as the notion of fluency discussed here, it is important to establish the basis on which speech is organised in the time domain. This point is particularly relevant, since most current prosodic models do not define this issue explicitly, and consequently, do not provide concrete temporal representations for speech (for examples in French, see Beaugendre, 1994; Di Cristo & Hirst, 1994; Jun & Fougeron, 1995; Mertens, 1990). As a result, the temporal structure of speech is most often reduced to the *accentual* structure of speech in these writings. We have shown that this underspecification leads to unacceptable deficiencies in areas such as high-quality speech synthesis.

Our point of departure is that speech is a human motor activity that shares a number of principles with other human motor activities, particularly those related to human communication (gestural, verbal, etc. activities). Similarly, we assume that the *basic principles of rhythm* are shared by a number of other motor activities associated with communicative activity (see *e.g.*, Grosjean & Lane, 1981; Evans *et al.*, 1986). We furthermore assume that any speech temporal model should observe these primary motor activity principles, since any spoken language structure is necessarily grounded in a motoric substratum.

As understanding of human behaviour can be improved through modelling, we shall consider the modelling requirements of this field. It will be shown that concepts and data on the temporal structure of speech fit quite well into a *chaos model*, which is an original contribution in this area. Finally, we will discuss how the chaotic nature of speech rhythm opens some interesting directions for language teaching.

A. The Basic Principles of Rhythm

1. Defining the Problem: Clocks vs. Memory Structures

The human tendency to produce and to perceive rhythm, *i.e.*, to organise several individual speech elements into regular groups, has been extensively documented by psychologists during the second half of the twentieth century (for a review of this question, see Evans *et al.*, 1986). Despite considerable work involving the precise measurements of temporal structures of speech as well as their symbolic and numeric modelling, some essential questions remain. For example, it is not clear exactly how an event is perceived as long or short, or as fast or slow. What is the time scale to which we refer in making such judgements? Or in terms of judgements of speech synthesis, how do we judge if an artificially produced sound is too long or too short? Furthermore, what is the temporal unit in speech perception[1]? Barbosa (1994), among others, assumes the existence of an internal clock, which could be the anchoring point for the production and the perception of speech rhythm. Is such a theory supported by empirical evidence?

In the following sections, the most relevant studies in psychology and biology on the issue of internal clocks will be recapitulated. This will identify a number of basic features of rhythm theory, which are of use in the analysis of the temporal structure of speech. Particularly, we will discuss the assumption that the temporal structure of speech is based on pulses provided by internal clocks. This will raise the more general question of the nature of the temporal unit in speech production and speech perception.

According to Gestalt theory, human behaviour can be guided by referring to a set of prototypical “forms” located in memory. A presumed temporal unit underlying speech could therefore be nothing more than a memorised proto-form abstracted from, and integrating into a whole, a large number of similar perceived units. These notions will be discussed in the light of recent results in neuro-physiological and auditory cognitive psychology, *i.e.*, psychoacoustics. It is important to delimit this putative temporal unit, since both perceptual coherence for heard speech, and rhythmic coherence of produced speech are assumed to be guided by such a unit — or a family of such units.

2. Temporal Perception and Inner Chronometry

2.1. Looking for a Perceptual Clock

Numerous studies have focused on the exploration of an hypothetical physiological mechanism capable of pacing a perceptual clock (Friedman, 1990). Among the examined candidates are, for example, cardiac and breathing rhythms, or alpha rhythms as captured from the scalp overlying

---

1 Taking as point of departure the principle that speakers are also listeners of their own speech, two subquestions of this point are under discussion. First, assuming a perceptual speech unit of a given duration, is such a unit equivalent to a temporally corresponding production unit (see "motor theory of perception", section 3 of this chapter)? Second and as questioned by Levelt (1989), can we even speak of a single temporal unit, or do we have to purport the psycholinguistic reality of a whole series of temporal units?

the cerebral cortex, or various autonomous cellular rhythms that can be expressed in terms of oxygen consumption, ion exchanges or genetic activity (for on overview on biological clocks, see Robert, 1989).

However, the existence of clock specifically serving to pace a "perceived temporal sense" could not be established for any animal or human being (Robert, 1989). According to Friedman (1990), there are many potential candidate clocks, but there is no possibility of clearly identifying one over another. One of the reasons for this difficulty is the fact that none of the candidates could be shown to maintain a continuous link with temporal perception. Moreover, psychologists have demonstrated numerous temporal distortions. For example:

- Interesting activities shorten the estimated passed time;
- A great number of activities performed within a given period lengthen the estimated passed time;
- A period of time is perceived longer if the subject knows in advance that the period will have to be estimated;
- The perception of time intervals in music depends partly on the tonal distance between sounds.

A further difficulty concerns the empirically documented complexity of internal clocks, which complicates any speculation about the relationship between the operation of a physiological clock, and highly complex human activities like speech. If such a clock existed, some link would have to be documented between the clock and at least one or several of the many processes triggered in speaking (ranging from motor intention to motor activation). Further problems concern the mode of interaction of various clock, since they do not all operate at the same speed. Moreover, the origin of such putative clock timing, as well as the way it is established and regulated, would have to be explained (Gregory, 1993). In other words, the relationship between such clocks and the temporal organisation of speech is neither documented, nor straightforward.

2.2. A Psychological Clock

According to Friedman's (1990) recapitulation of psychological studies of the preceding thirty years, many human behaviours relating to time estimation, such as distortions, are better explained by ordinary mental processes than by internal physiological clocks. In this view, the temporal estimation of a period is related to the density of mental processes (sensations, ideas, feelings) occurring during this period. In that sense, biological rhythms, with their potential afferent inputs into the CNS, are only one of several sources that the neurocortical system may exploit to establish temporal structure for motor events.

From a psychological point of view, timing may cover various realities. It can be a stream of internal and external events, a network of locations in memory, patterns built into our representations, or co-ordinates to orient oneself in the present. Robert (1989) and Friedman (1990) suggest that a human being's temporal universe is based on a personal mental construction process. It is an adaptation of one's general reference systems, acquired in an extensive interaction with the multitude of relationships established between external experiential events and internal bio-psychological clock events.

This complex interdependency provides meaning and dynamic orientation to our activities. Finally, this inner construction is also based on our social experience (culture, common experiences, etc), and it is what makes it possible to partly share one's temporal universe with others. From calendars or watches to social living conventions, many external temporal frames regulate social interactions. Some of the effects of this social constraint on oral communication will be demonstrated in a subsequent section.

For the time being, the hypothesis of a complex relation between bio-psychological multi-temporal processes and the basis of speech temporal organisation is thus the most likely explanatory hypothesis. As will be seen, this assumption will be essential for postulating a generally valid theory of speech rhythm.

3. Pace Regularity as Global Shape Perception

In psychoacoustic experiments conducted by Bregman and others (*e.g.*, Bregman 1994), it has been shown that grouping principles, originally described by Gestalt theorists for visual perception events, also apply to sound perception (*e.g.*, Patel *et al.*, 1998). The perceptual unit for the recognition of an object emerging from this theoretical framework would seem to correspond more closely to a global "shape" — *i.e.*, a certain structure of features — than to an "atomic analysis" of each constituent element of the perceptual object, as originally suggested by Helmholtz[2].

The well-attested existence of a capacity for grouping acoustic stimuli on the basis of similarity and temporal coherence can be linked to the older notion of *isochronism* in speech, to be discussed more extensively below (Dauer, 1983; Evans, 1986; Lehiste, 1977). In that view, certain elements of speech (typically stresses, accents or syllables) were thought to occur with a certain regularity, and any deviation from the main clock rate would suggest the presence of a special event (linguistic or whatever). Furthermore, such a grouping activity (*i.e.*, the construction of rhythm) was traditionally assumed

2 Warren (1994) suggests that the two perceptual modes co-exist: a synthetic, global mode which acts as the "default mode", and an analytic mode which permits an identification of the individual tones making up the group.

to require a benchmark, which in turn was traditionally linked to the putative pulsation of an inner clock.

Despite its intuitive appeal, most elements constituting the isochronistic hypothesis have been refuted. Isochronistic markers have proved elusive, since neither stress in so-called stress-timed languages, nor syllable length in so-called syllable-timed languages show the expected measurable regularity (see references below). Furthermore, in view of the above discussion on the improbability of a simple clocking mechanism, the underlying mechanism of whatever "isochronism" is perceived would probably have to be restated in terms of the construction and use of a complex and interactional cognitive organisation serving temporal recognition and the temporal structuring of actions. Indeed, a reformulated and tempered notion of *perceptual grouping of temporally similar events* permits a link to the empirically demonstrated grouping principles subserving global temporal global shape recognition.

This, in turn, introduces the next major issue, which is the problem of the underlying perceptual and cognitive mechanism for temporal shape recognition and processing. Assuming that global shape perception serves as the basis for the perceived pace regularity in speech, how does it work? One can begin to answer this question by asking whether there are any differentiations between afferent and efferent use of temporal shape concepts.

According to Gestaltist shape theory, an essential distinction is made between perception and action. In this framework, *perception* establishes links to a set of reference points, that is, to a set of *durable and invariant prototypical forms* stored in memory, while *action* employs such forms to guide behaviour. However, Gregory (1993) has noted that neurophysiological evidence does not support such an explicit distinction between afferent and efferent use of prototypical forms. And indeed, a consultation of experiments listed in the Medline database[3] — specifically abstracts of relevant studies in psychophysics, psychoacoustics and neurophysiology from 1993 to 1998 — would tend to support his observation. The Gestaltist model by itself cannot account for how neural structures perform complex perceptual tasks of integrating large sets of data into a coherent whole, it merely poses the problem.

Gregory (1993) gives the example of perceiving a very large object, which imposes a rapid eye scan of various points of the object. In this task, different pieces of the object are captured and globally interpreted in a very short time. Gestaltism by itself cannot explain how perceptual coherence is

---

3 Medline; the National Library of Medicine's Journal archive developed under the National Institute of Health.

rapidly constructed from such multiple inputs. Moreover, as generally stated, Gestalt theory does not consider that human beings perceive their world with multiple various sensors: vision, audition, touch, etc. This has led some researchers, among them Berthoz (1997), to suggest considering perception as an *active multimodal process:*

> «The issue of perceptual coherence is not only a geometric or a dynamic problem. It also assumes active central mechanisms to allow disambiguation, compensations or anticipations of differential delays between sensors, to allow the unification, by means of astute biological mechanisms, of sensory spaces that are not only changes in co-ordinates.» (Berthoz, translation, page 101)[4]

Berthoz suggests that instead of "representations" of durable and invariant prototypical forms, perception should be conceived as simulated actions projected out in the world. This rejoins the motor theory of perception (Liberman & Studdert-Kennedy, 1978; Liberman & Mattingly, 1985), which says, in the main, that speakers can only hear what they can articulate, thanks to their internal representations of speech. While the motor theory of perception finds some support in empirical studies in neuropsychology (Damasio, 1997), recent studies on infant language acquisition also show that the precise formulation of the relationship between perception and production needs some adjustments, since relations between language perception and articulatory performances are not entirely straight-forward. This is because certain stages of language acquisition trigger reorganisations in infant capacities, and during those periods, infants seem to regress with respect to articulatory skills, without showing any parallel perceptual deficits (Boisson-Bardies, 1999; Konopczynski, 1994). At certain stages of learning, perceptual capacities can thus surpass articulatory capacities.

Overall however, it seems safe to conclude that global shape abstraction for temporal events in speech results from a complex interaction between perceptual and production events, in which productive experimentation with speech events during the acquisitional phase, both individually and socially, plays an important role. This would account for the results of a large number of neurobiological, psycholinguistic and acquisitional studies (*e.g.*, Cutler, 1994). It is also coherent with studies showing that the temporal benchmark may vary, *i.e.*, that the temporal reference unit used in rhythm perception is variable. Sendlmeier (1995) suggests, for example, that a listener is capable of changing his temporal window in speech perception. It is clear that none of these studies offer direct support for the hypothesis that an inner clock

4 «Le problème de la cohérence (perceptive) n'est pas seulement un problème de géométrie ou de dynamique. Il suppose des mécanismes centraux actifs qui vont permettre la levée des ambiguïtés, le rattrapage ou l'anticipation des retards différentiels entre capteurs, l'unification des espaces (sensoriels) par des mécanismes biologiques astucieux qui ne sont pas seulement des changements de coordonnées, etc.» (Berthoz, page 101).

provide an inner, durable, and invariant pulse. Finally, they also emphasise the importance of an adequate temporal coordination between all components involved in speaking.

After this consideration of inner clocks and pace regularity, consequences for a theory of speech temporal organisation can be discussed.

4. Speech Rhythm

As a preliminary to a discussion of the theory of rhythm theory, it is important to circumscribe what the notion of "rhythm" may encompass according to various authors and studies. Here is a short list[5] for the item "rhythm", established on the basis of various studies on speech:

Succession of beats
Stimulus sequence organisation
Stimulus sequence structure
Stimulus sequence structure, accent excluded
Accent distribution
Accent repetition
Regular accent repetition
Accent repetition giving the impression of regularly repeating passages
Accents and pauses over a given period
Alternation of accented and unaccented syllables
Equal number of unaccented syllables
Equal number of rhythmic groups
Unequal syllables
Number of syllables per rhythmic groups
Accents, pauses, tonal variations, breaks and syllable durations
Phrases-like units
Various speech rates
Tempo

The definition given in the well-known French dictionary *Le Grand Robert* is also rather vague:

«A essential harmonic feature, which differenciates poetry from prose by imposing on the text a regular distribution of strong beats, accents and breaks, number of syllables, etc.»

Similarly, Evans *et al.* (1986) have shown that rhythm may refer to or may involve:

time,
time and space,

---

5 Grateful acknowledgement goes to Mme Odile Ledru-Menot, Professeur en Didactique des langues at Paris 3 University, for her contributions to this inventory.

dynamics in time,
musical structure,
individual perceptual capacities
motor action in response to an outer source,
capacity to determine differences,
an esthetic quality of movement

This prolixity of meanings for a single item facilitates neither the accessibility nor the construction of a coherent theory. Moreover, the multiplicity of meanings for rhythm associated with unquestioned uses of the item may also lead us to conclude that there is agreement or disagreement in the scientific community where there is in fact none. Since it is also difficult to avoid entirely the use of this term, we shall retain Crystal's definition[6] as given in his dictionary, that is, "rhythm is the regular perception of prominent units in speech". Speech rhythm is thus understood as a cognitive construction based on a body of features, which may change with time and situation.

4.1. Fundamentals of the Temporal Structure of Speech

According to the phoneticians Di Cristo & Hirst (1994), who work in Aix-en-Provence, France, the perception of rhythm is based on a combination of temporal structure and syllabic prominence (in the temporo-frequential domain). Fraisse (1974), a French psychologist, considers that temporal factors are primary in the perception of rhythm: durations are assumed to be more significant than accents. However, Grover & Terken (1995), two Dutch phoneticians, have presented the opposite view by suggesting that accent is primary in providing a rhythmic impression. At this point, it is difficult to tell whether these discrepancies are due to the subjects' native language, or to the research paradigm.

For the rest, the prosodic literature does not show a systematic effort to approach the temporal component independently of the intonational component. As summarised in the next section, *the temporal structure of speech* has largely been approached through the linguistic filter of accentual structures. In view of the importance of the coordination and integration of a wide range of temporal perceptions (arguments developed in Section 3), this is regrettable. It seems to us that the "accentual filter" imposed on typical phonetic analyses by linguistic theoretical preconceptions poses a hindrance to a much richer set of analyses, analyses that could ultimately lead us to a much more profound understanding of how speech rhythm is constructed and perceived.

---

6 Crystal D. (1990). A Dictionary of Linguistics and Phonetics. 3rd Edition. Blackwell Ed. London.

- *Temporal Regularity*

There is a distinction, initially introduced by Pike (1945), between stress and syllable paced languages. Stress-paced languages like English or German are characterised by relatively great variations of syllabic durations, and by a diversity of syllabic structures. Conversely, syllable-paced languages such as French or Spanish are characterised by smaller variations of durations and by a greater regularity in syllabic structures. In part, this is due to the high frequency of open syllables in such languages (Goldman *et al.*, 1994), which reinforces the impression of temporal regularity between syllables (Dauer, 1983). Dauer (1987) showed, for example, that in addition to the high frequency of CV and CVC syllabic structures in French, active historical mechanisms favour the more regular syllables, thanks to the simplification of clusters in final position, to epenthesis and to liaison, all of which have the net effect of preventing the creation of deviant syllable structures.

Beyond these phonological considerations, Keller & Zellner (1996) have shown that in French, a major contributor to the impression of rhythmicity arises from the segmental level alone, independent of any syllable or phrasing effect. In their tripartite durational predictive model for French, the segmental level alone explains a full 48% of total variance of segmental duration. This means that proximal segmental information (*i.e.*, which segment precedes, which follows, and which is to be pronounced) captures a considerable portion of the entire durational structure, and synthetic speech produced with durations calculated on the basis of segmental information alone exhibits some surprisingly "rhythmic" characteristics[7]. Furthermore, current studies in our laboratory confirm that for stress-paced languages such as German[8], the importance of the segmental level is equally considerable. First estimations suggest that around 58% of the durational variance of segments is explained at the segmental level alone. This implies that the sound organisation along the segmental stream constitutes in itself a very powerful parameter for the construction of French or German temporal structure of speech.

Having said that, it remains true that this information is clearly insufficient to provide the whole auditory perception underlying natural temporal coherence. This also emerges from the segmentally calculated synthetic speech, where accelerations and decelerations marking phrasal groups are painfully absent. Additional parameters clearly need to be incorporated at the syllabic and phrasing levels for full rhythmic speech to emerge.

---

7 The interested reader can perform the necessary experiments with our freely available speech synthesiser LAIPTTS, available at http://www.unil.ch/imm/docs/LAIP/LAIPTTS.html.
8 This study has been conducted by Dr. Siebenhaar in our laboratory.

• *Segregation and grouping*

At the next higher level, syllables are usually joined into larger groups, though exactly what the size of that group may be is open to some discussion. Some authors suggest that the basic grouping principle is the final group accent which serves to identify and delimit the "phrasal" or "accentual" group (e.g., Bailly, 1989; Di Cristo & Hirst, 1994; Ladd & Campbell, 1991; Mertens, 1990). However, the fundamental grouping principle at this level may also be the repetition of the initial group accent in certain languages, or the repetition of certain sounds, syllables, or syntactic marks. For example, Barbosa (1994) considers that the rhythmic basis for French is the period included between two “Perceptual-Centers”, or "P-centers"[9], a duration that has been shown to correspond more or less to the period located between two vowel onsets. Delais (1994, 1995) and Pasdeloup (1992a, 1992b), also working on French, assume that the rhythmic period is equivalent to the accentual group, that is, the interval included between two accented syllables.

In addition to the problem of identifying the rhythmic unit, the unit itself may change according to the situation. Guaïtella (1992, 1996) and Zellner (1997, 1998) suggest in their studies on French that rhythmic structures should be differentiated according to the style of speech. For example, a neutral reading task performed in uninterrupted and quiet conditions poses few respiratory constraints, and it limits a number of further communication constraints, leaving the speaker to be guided essentially by the text's graphic structure (Guaïtella, 1992). In that case, reading conforms largely to a metrical temporal organisation, *i.e.*, an easily predicted organisation, with more or less equivalent time periods. One corollary of this is that neutral, declarative speech is relatively easy to implement in speech synthesis. It also permits Wang & Hirshberg (1992), among others, to use textual features in addition to intonational criteria in their prediction of prosodic boundaries.

Conversely, in the case of a spontaneous speech sequence, the temporal structure is far more complex. The temporal structure not only has to continuously adapt to situationally changing constraints from the respiratory point of view, but also in terms of interactions with communication partners. Under these circumstances, rhythmicity has been shown to be more directly anchored in sound contrasts (Guaïtella, 1996).

In conclusion, there is a common agreement concerning the verbal stream as structured into rhythmic entities. It is universally conceived of as a perceptually prominent, repetitive event. However, there are relatively complex discussions concerning the precise identification of the different

---

9 For a review of the question of the P-Center ou “Perceptual-Center”, please consult Barbosa (1994) and Scott (1993)

rhythmic units. More clarity on the issue may emerge from examining the next question, which is how this temporal space might be traversed. In other words, how do we calculate an adequate speech rate to pass from one rhythmic entity to the next? For example for French, syllabic durations will materialise this temporal course. We assume that the way the rhythmic space is traversed is also an important issue in the perception of temporal coherence.

4.2. The Isochrony Illusion

"Isochrony" is a hypothesis of speech metricity, in which speech is supposed to be marked by rhythmic benchmarks separated by regular periods. This principle involves guidance by an "inner clock" (McAdams & Bigand, 1994), the scientific basis for which has already been discussed in this article. Since the notion of "isochrony" has inspired many phonetic studies on rhythm, we shall now turn to a more detailed examination of this principle.

Allen (1975) noticed that the human ear tends to perceive period sequences much more regularly than they are in effect. Generally, long periods are underestimated and short periods are overestimated, resulting in a tendency towards "isochrony". Isochrony is thus chiefly a perceptual effect, rather than a speech production event (see also Lehiste, 1972, 1977). In that sense, Fónagy (1992) noticed that the tendency for chronometric harmonisation of accentual groups in Hungarian hardly ever ends in the *full equalisation* of durations, even though a partial equalising trend can be demonstrated. Wightman (1992) confirms that he found no evidence for isochrony in his data on English.

For those who do assume at least some "isochrony" in speech, observed deviations from the metric principle are diversely explained. Barbosa (1994), who works on French using an inferred (*i.e.*, not directly-measured) psycho-acoustic mark of isochronous speech perception, the "P-Center" or “Perceptual-Center”[9], considers that the listener *expects* temporal regularity. When disruptions in the metric principle do occur, they are interpreted as a speaker strategy to alert the listener to special events. This view is not shared by Duez & Nishinuma (1985), also working on French, who explain irregularities in rhythm as a “standard pattern”, according to the "alternation principle", as formulated by Woodrow in 1951 (see Duez, 1985). Duez and Nishinuma's data first contradict any isosyllabic tendency (*i.e.*, they do not find regular syllabic durations), and second, they suggest two trends for French: relative brevity of the initial syllable and duration alternation for succeeding syllables. Similarly, Wenk & Wiolland (1982) and Vaissière (1991) show the importance of syllable lengthening in French, as well as the major role that syllable lengthening plays from the perceptual point of view.

Another argument against isochrony arises from the predictability of accelerations and decelerations in speech. Caelen-Haumont (1991), Pasdeloup (1990, 1992.b), and Keller *et al.* (1993) show that in French reading tasks, the utterance is temporally organised into various slowing sequences. Typically, within an elementary temporal sequence, the duration of a non-accented syllable is lengthened as a function of its proximity to an accented syllable (see also Nishinuma & Duez, 1988; Nishinuma & Santi, 1992, Vincent *et al.*, 1995). Subsequently, each new basic temporal sequence is characterised by a sort of “reset” of the syllabic duration.

It is interesting to speculate on the origins of these de- and accelerations. While some of the systematic deviations from expected durations (e.g., durations calculated on the basis of segmental information alone, as above) can be related to fairly evident phrasal grouping principles (*e.g.*, slow-downs at the end of phrasal groups), some of the more subtle de- and acceleration effects may also have to do with acoustico-articulatory constraints. In this context, it is interesting to consider some principles arising from the study of acoustico-articulatory relationships, notably the concept of "cyclical reference", which has been reintroduced on the basis of quantum[10] tendencies in articulation (Stevens, 1989; see also Boë et al, 1992, for a review of the theory).

Stevens divided the acoustico-articulatory space into three regions (figure 1). Regions I and III are envisaged as stable acoustic regions, while region II is considered to be a transitional, unstable zone. A typical example for the first type of acoustico-articulatory space is the English /a/-vowel space, where a wide variety of pronunciations of the vowel are easily perceived as members of the /a/-space. The second type of space is exemplified in English by the /s/-/S/ distinction, where only minor tongue readjustments result in a perceptual differentiation. Concretely, this means that phones produced in regions I and III need less articulatory accuracy to be perceived correctly than those produced in region II, since the acoustic effect of an inaccuracy is considered to be insignificant for speech communication. In region II, relatively minor articulatory variations provoke perceptually relevant acoustic changes, which in turn imposes a delicate articulatory organisation on the concerned vocal tract region. This vocal tract region is named the quantum region. (A new interpretation of the quantum region will be suggested in the second part of this article.)

---

10 Quantal tendencies, in the sense that there is dependence on numbers expressing limited quantities (i.e., “quanta”). For a discussion of the quantal theory of speech, see Stevens K. (1989, On the Quantal Nature of Speech. Journal of Phonetics, 17. 3-45), and succeeding articles.

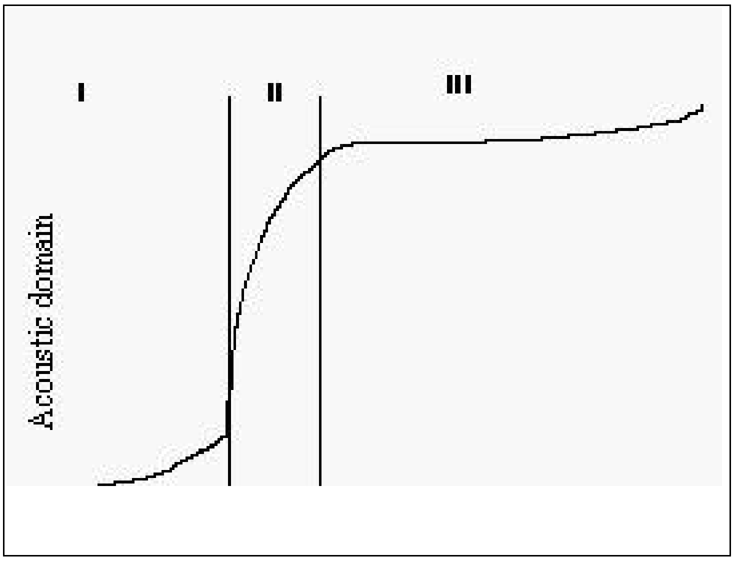

Figure 1. Acoustico-articulatory space as defined by Stevens (1989).

This quantum principle may well be relevant to speech rhythm, since Fant *et al.* (1996) postulate the same kind of quantum feature for the temporal organisation of speech[11]. They assume the existence of some rather sensitive temporal quanta. For example, they suggest that durations of segments, unaccented syllables, accented syllables, and accented syllables located in the vicinity of a syntactic boundary are all related by even-numbered factors: 1:2:4:8. According to this hypothesis, durations falling into the general domain of such relations are considered to be "easy perceptual targets", while durations falling into the border zones between durations calculated on such factors would be perceptually more delicate. In this manner, durational and rhythmic expectations are built up that provide a temporal framework for both speaker and listener.

In summary, speech production involves a number of progressively more complex, dynamic processes. A natural speaker's speech rate is hardly ever steady over several utterances. A speech timing theory must take this variability into account. For this reason, it is difficult for us to support the inference of the existence of a continuous, explicit and stable relation between speech periodicity and supposed regular beats originating from inner clocks, as suggested in Barbosa's thesis (1994). We lean more heavily towards the interpretation of the psychoacoustician McAdams (1994), who

[11] While certainly provocative, Fant's proposal needs further verification. In our data, we could not verify some of the consequences arising from this proposal (Zellner, 1998).

suggests that the temporal structure of speech is like a set of frames, that is, like a set of complex articulo-perceptual schemes established during language acquisition. Similarly, his colleague Warren (1994) considers that different integrative mechanisms may be used during speech processing. These conceptions, in turn, permit close links to Levelt's psycholinguistic hypothesis (1989) that speakers use various processing units of different sizes, given the circumstance. In that sense, speech temporal organisation results from a complex relationship between numerous bio-psychological processes, each of which imposes its own time frame, as will be shown in the following section. It seems to us that any successful speech temporal model would have to be founded on an understanding of this complex relationship.

5. Speech Fluency: The Management of Temporal Constraints

In previous sections, we discussed the weakness of hypothesising an inner clock as being “the” essential timing supervisor for a supposed "metrical speech rhythm". In this section, it will be shown that our hypothesis of a bio-psycho-sociological organisational structure (Zellner, 1997) can provide a viable alternative framework for the characterisation of rhythm, formulated against the background of major currents of empirical knowledge on the temporal structure of speech.

We issue from the assumption that a three-level set of constraints effectively limits speech activity in the time domain: bio-psychological, social and pragmatic constraints. Speech temporal organisation involves the integration of, and compromises between *all* of these pressures, which evidently exceed the sole linguistic level. The construction of the temporal framework for a given utterance is seen as an exercise of optimisation between these constraints, in other words, as an *emergence of an optimised temporal structure appropriate to a given set of circumstances*.

5.1. Not Everything is Possible

As speaking is a complex dynamic process involving the integration and the synchronisation of various psycholinguistic, linguistic, motor and perceptual processes, an important issue is to understand how this synchronisation is made possible. As a physio-psychological and linguistic event, fluent speech may be interrupted by the effects of three components of human functioning, bio-psychological constraints, sociolinguistic factors and pragmatic conditions (Zellner, 1997). These three components participate in speech fluency by creating a *temporal frame*.

More in detail, speaking is first limited by bio-psychological constraints, such as speech motor production and psychological processing constraints. For example, any phonatory activity is evidently circumscribed by respiratory needs. Also, speech cannot be accelerated or slowed beyond certain bounds, partly because of the intrinsic duration of each phonatory gesture, and because of inescapable delays in perceptual processing (Zwicker & Feldkeller, 1981; Botte *et al.*, 1989). Moreover, since some segmental durations are longer than others, some phonatory gestures have to be delayed, while others have to be anticipated. At the acoustic level, this complex set of motor commands is concretised in a sound stream sprinkled with various silences, and also, with acceleration and deceleration periods. This is one fundamental reason that renders fluent speech discontinuous.

Many studies in psychology and experimental psycholinguistics have shown that discontinuities along the sound stream do not occur just anywhere. For example, in French spontaneous speech, Zellner (1992) found that hesitating periods (typically, “euh” and false departures, repetitions, sound pauses) occur at some very specific moments in the utterance. For instance, normal speakers — *i.e.*, those with no discernible speech pathology — never hesitate in the middle of a syllable production, and they hardly ever hesitate during a lexical word production. More often, they hesitate between two meaning groups. These discontinuities, with or without hesitation, contribute to the perception of a certain cohesion between words (Goldman-Eisler, 1968, 1972; Boomer & Ditman, 1962; Cook, Smith & Lalljee, 1974; Chafe, 1980, Grosjean & Deschamps, 1975). In that sense, a pause may be considered as the expression of cognitive processes underlying speech.

Speech temporal organisation is also restricted by sociolinguistic habits. First, each language has its own sound and prosodic systems. For instance, one linguistic (dialectal) characteristic concerns the duration of sounds. Typical English durations are different from typical French durations, and typical French Swiss durations are again different from typical Parisian French durations. Second, beyond linguistic systems, social-cultural standards like speech rate add another temporal pressure on speaker’s performances (Miller *et al.*, 1976; Street & Brady, 1982). For example, Street and Brady show that a socially attractive effect is induced when speaker’s speech rate approximates an interlocutor’s speech rate. Moreover, the interaction established between speakers may influence their prosodic patterns, due to a desire for identification or a social integration effort. This kind of temporal adaptation is currently described in social accommodation theory[12].

Finally, the temporal dimension of speech is also affected by pragmatic and affective events. For example, for Caelen-Haumont (1991), the way a speaker passes from one meaning group to another, and the way a speaker

12 Accommodation theory: see e.g., Newmeyer (1988, Vol IV, chapter 6).

builds cohesion between words in a French reading task[13] depends on text meaning, on the speaker's intentions and on his perception of the situation. Also, a particular affective intention may change the temporal dynamics — at least temporarily, due to changes in tension and elasticity of the vocal folds or changes in the respiratory rhythm (Scherer, 1984, 1986, 1989).

5.2. Integration of Actions in Speech

Speaking thus involves the taking into account of these three constraint ranges and underlying processes — psycholinguistics, linguistics, articulation — since they all have their own temporal specifics. In that sense, fluency designates the capacity to coordinate various processes operating at different speeds.

For example, if we suppose for the moment that the underlying processing operations are not correctly synchronised (e.g., because of a delay in the message conception, or a delay during the lexical access, or an inadequate motor timing plan), this may generate temporal disruptions. Temporal disruptions, *i.e.,* dysfluencies, are not only induced by difficulties located at the bio-psychological level. They may also happen in the context of second-language learning, for example. In that case, the speaker may use a strange speech rhythm, possibly with many hesitations. Speaking in a difficult communicative context is another typical situation where dysfluences may occur: for instance, while speaking and verifying information on a computer at the same time, or when forced to speak despite the influence of a disturbing emotion. Conversely, the capacity to talk fluently at a rapid speech rate reflects a speech activity that is trouble-free at all three constraint levels.

Certain aspects of this theoretical framework for temporal organisation are well-understood and well-documented. However, one crucial question still awaits an answer: what are the principles governing the attainment of an optimal compromise structure for speech timing, given the various constraints acting on the production of the speech utterance? We shall turn to chaos theory, a set of mathematical concepts and tools that may well throw some further light on this question. We shall see that it may be useful to examine speech timing from this angle, since it points up several characteristics that are typically associated with chaotic evolutions: sensitivity to initial conditions, partial periodicities and partially unpredictable trajectories

13 Notice that, although reading is a particular type of speech, Gitta and Laan (1992) could not establish clear differences between the two types of speech in perceptual tests.

B. Speech Rhythm through the Light of Chaos

1. What is Chaos?

In physics, chaos is not equivalent to total disorder. A non-linear dynamic system is chaotic when the evolution of its key parameters, although perfectly deterministic (*i.e.*, non-random), are very difficult, or close to impossible, to predict. A nice definition-in-a-nutshell of chaos is: “Chaos theory is the qualitative study of unstable aperiodic behaviour in deterministic non-linear dynamical systems” (p1, Kellert, 1993). It will be our task here to take these elements individually, and to relate them to various aspects of speech timing and speech behaviour. We shall first describe some of the characteristics of typical chaotic systems in some down-to-earth terms. In the following sections, we shall relate these characteristics to speech timing.

For chaotic behaviour to appear in the first place, a system must show *serial dependence*. This means that a system's exact state at a given time is either partly or entirely dependent on its previous state. It is this serial dependence that leads to the chaotic behaviour of a system: since the same serial effects cumulate over time, exceptional states as well as abrupt reactions to these exceptional states can be induced, which ultimately results in the rather variable trajectories that characterise chaotic systems.

A second typical criterion is a system's *complexity*. Most often-cited examples of chaotic systems are complex: weather systems, turbulence in gases and fluids, human cardiac functions, etc.

Another characteristic, directly related to the previous two, is a *high sensitivity to initial conditions*. A complex system subject to a multitude of external conditions will often show long periods of rather stable conditions. This is because large numbers of input parameters have ample opportunity of cancelling out each other's effects. But sometimes, particularly under the influence of a particular set of cumulative serial effects, unusual conditions may occur. This is the proverbial case where a tiny change in the initial data of the system triggers rapidly dramatic changes. A famous example to illustrate this mathematical property is the “butterfly effect”, where the proverbial "flap of the wing of a butterfly in Rio can provoke a snowstorm in New York".

While chaotic systems are subject to important *limitations in the predictability* of their evolution, their *occasional convulsions* may also elucidate some unexpected causal relations. In that sense, chaotic events can throw some interesting new light on the dialectic between order and disorder. Also, one may question to what extent human functioning is *deterministic*, *i.e.*, non-random, and if it is deterministic, to what extent it is *chaotic*.

In recent years, several aspects of human behaviour have been modelled as a non-linear dynamic system. For example, a number of studies have examined walking, in order to determine if the stride interval of the human gait is random. Many models have problems reproducing the observed behaviour correctly. In a recent study, Hausdorff, J.M. *et al.* (1995) established long-range correlations in walking: the stride interval at any time depends on the stride interval at remote previous times. This kind of sequential ordering and serial dependence is hard to predict, however it *is* deterministic, and thus the system can be said to be chaotic.

Below, we will shall see that speech rhythm presents some characteristic properties of a chaotic system. This knowledge permits us to estimate the limits of predictability for a temporal model of speech behaviour. If a system is deterministic, in theory and with sufficient informational input, its future behaviour is knowable. If it is also chaotic, we know that we should not only consider *stable predictors* in making durational prections, but also *dynamically evolving and possibly cumulating* effects. The predictions of current models in speech synthesis, for example, are all unsatisfactory to some degree, since they are incapable of explaining 100% of the durational variance. Such timing models use predominantly stable inputs (information about segments, syntactic structures, etc.). Cumulative serial effects or longer-term serial effects are generally left aside. Concretely, this means that any predictive model for speech timing generates some noise. If we assume with chaos theory that this "noise" is not random, but simply unexplained, it may well be that some of these residuals could be eliminated.

2. Speech as a Chaotic System

The typical characteristics listed above for chaotic systems find their counterparts in speech behaviour and speech time. As we have seen in this review, speech timing (as speech behaviour in general) shows *serial dependence*, since speech timing is dynamic (it evolves over time), and certain durational changes can only occur in the context of certain preceding states. This is useful, because it lets us postulate that certain timing effects are directly dependent on previous temporal states, that is, it lets us put "coherent temporal shape perception" into the context of chaos theory.

Also like other typical chaotic systems, speech and speech timing is *complex*, in the sense that it is subject to a great number of influences, some of which may be linear and some non-linear.

Another property of chaotic systems, their *sensitivity to initial conditions*, also has counterparts in speech. Small variations in initial conditions may indeed induce dramatic changes later. Steven's and Fant's proposals on speech quanta suggest that minor articulatory variations in certain regions of the vocal tract trigger major changes in acoustic parameters. This configuration imposes high articulatory accuracy in this region of discontinuities, which is the quantum region.

Another example of this sensitivity concerns the production of speech rhythm, where initial conditions can be seen to correspond to the initiation of the motor plan giving rise to a number of possible outcomes. A number of studies on speech stuttering and speech dysfluencies have shown that in certain circumstances, a tiny delay in motor synchronisation may induce major speech dysfluencies a few words later in the utterance (*e.g.*, Rosenfield, 1989; Starkweather, 1987). Under other circumstances, small disruptions do not impede speech fluency. Empirical knowledge acquired in the speech synthesis community establishes the same kind of property for the modelling of timing: some regions within the speech stream are very sensitive to any temporal changes and impose high accuracy, while other regions are show more robustness. This was observed at the phasing level, as well as at the word, syllabic and segmental levels. A more extensive review of this question will be presented in a forthcoming paper.

Altogether, there are sufficient initial similarities between typical chaotic systems and speech behaviour to examine the question of the chaotic nature of speech timing more closely. We will do so with the help of a number of graphing and analysis methods.

3. Phase Plots and Hurst Coefficients

The phase space is used to represent the repetitive behaviour of dynamic systems. Dynamic systems are characterised by two components: their state at a certain moment, and their dynamics, *i.e.*, their evolution over time. The phase plot represents the evolution of a system, and the coordinates of the plot generally correspond to the system's components, *i.e.*, the number of degrees of freedom that characterise the system. The evolution of a movement can thus be represented by the intersection of its position (a point in the plan) and its velocity (the displacement of this point).

When a periodic system moves like an ideal pendulum (where no energy is lost and all oscillations are always identical), the phase plots shows a regular circular shape (cf. figure 2).

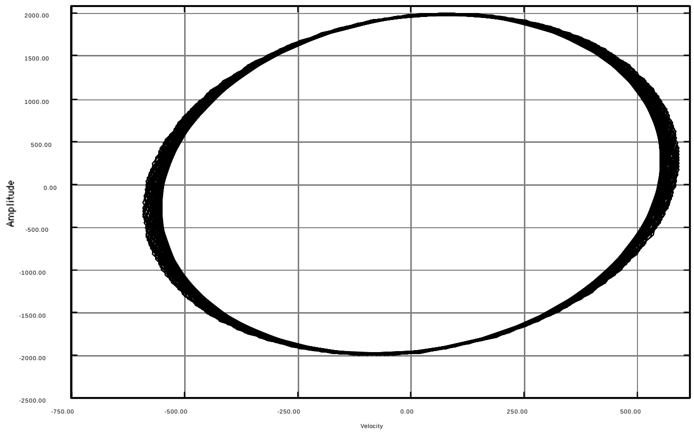

Figure 2. Phase plot of a periodic (sine-wave) signal

When a system evolves like a real pendulum, the amplitude of oscillations become smaller and smaller (because of friction), and finally, the system reaches a stable central point. This central point is called an *attractor*, a characteristic of this type of system. Other types of attractors have been identified for other types of dynamic systems. In quasi-periodic systems, several independent oscillations may co-operate, producing several toric attractors in the phase plot. Other complex and unpredictable systems tend towards “strange” (less precisely defined) attractors. On the other hand, *random* dynamic systems *do not* evolve towards attractors, because trajectories are bordered on surfaces of equal energy (Crutchfield *et al.*, 1989).

Recently, in many areas, researchers have discovered that many “random” phenomena were in reality associated with strange attractors, in particular heart beats and neural electrical activity (Crutchfield *et al.*, 1989). We examined phase plots of syllabic durations for fast and slow speech rates, and compared these plots with random values. The syllabic durations were extracted from a French "neutral" reading of 50 sentences at two speech rates and the phase plots represent an initial period of 2000 durations extracted from the initial portion of the recording session (cf. figure 3).

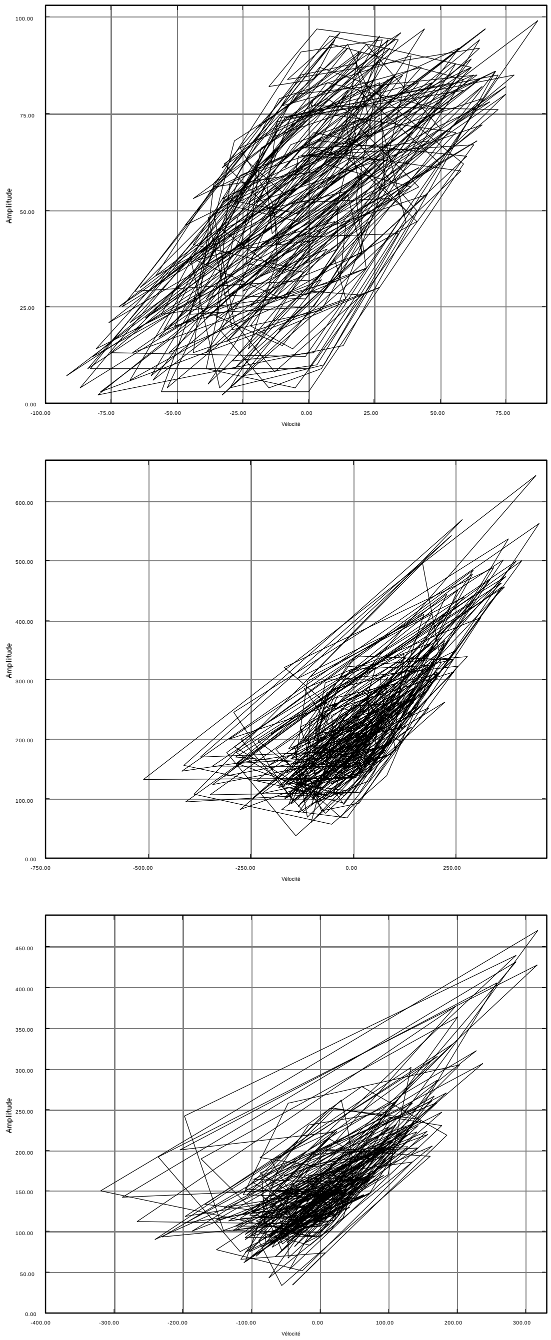

Figure 3. Phase plots for pseudo-random durations generated (top), syllable durations produced at slow speech rate (centre), and at fast speech rate (bottom). An initial series of 1500 syllables were extracted from the beginning of each recording session. Any other period would have produced the same figure. The two phase plots representing syllabic durations both converge towards an attractor, but the phase plot representing random values shows no particular tendency.

Syllabic durations phases show clearly how trajectories both converge towards three “strange” attractors. The trajectories between the three attractors are unpredictable, that is, they do not revolve always in the same sens. In effect, the trajectories in both cases are not regular circles, confirming that French rhythm is a not a simple periodic oscillation. The non-randomness of these time trajectories was also examined in exploratory fashion by examining the Hurst exponent[14] using the program Benoit 1.2[15] (Please note that contrary to usual practice, the Benoit programme sets the Hurst exponent for random motion to zero). It was found that the series of slow and fast syllable durations were systematically divergent from the results of the pseudo-random duration series (for reasons explained in the footnote, no significance tests are applied to these figures). This is a further demonstration that non-random effects are "hiding" in the series of syllable durations, effects that may well contribute at some future date to the improvement of the prediction of speech timing and speech rhythm.

	N	Windows	**H-exp**	SD	x
fast	982	22	**0.0989**	0.024386	2.15
slow	1050	20	**0.0334**	0.01966	2.51
random	982	22	**0.00812**	0.02121	2.52

Table 1. Results of a Hurst analysis using Benoit 1.2 on syllabic durations for two speech rates and random values. Contrary to usual practice, the program Benoit sets the Hurst exponent for random series to zero instead of 0.50.

We can conclude that sensitivity to initial conditions, phase plots and serial tendencies give excellent support to the hypothesis that speech timing and speech rhythm are chaotic systems.

4. Towards a Reinterpretation of Speech Rhythm and Future Directions for Teaching Languages

In the previous section, the chaotic nature of speech rhythm was established for French syllabic durations (1050 and 982 syllables), which were extracted from a neutral reading corpus. In a forthcoming paper, we will demonstrate more extensively the chaotic nature of speech timing in several languages.

---

14 Hurst Exponent(H): A measure of the bias in random motion. H=0.50 for Brownian (totally random) motion. 0.50<H<1.00 for persistent, or trend-reinforcing series. 0<H<0.50 for an anti-persistent, or mean-reverting system. H corresponds to the slope *b* for the relationship between log(time period)/log(average of rescaled ranges at that time period). The inverse of the Hurst exponent is equal to alpha, the characteristic exponent for Stable Paretian distributions. The fractal dimension of a time series, D, is equivalent to 2-H. Attention: Recent Monte-Carlo calculations of H performed in our laboratory have shown that the distribution of H is much wider than standard values calculated for slope *b*. Consequently, no significance values are given in the text.

15 Benoit 1.2.: Fractal Analysis System. TruSoft International. Inc., St. Petersburg, FL, USA.

The aim of the present article was to provide a general background for the understanding of speech timing, fluency and rhythm, against the background of acquisitional patterns of second language learning.

Some principles are worthwhile to retain, in order to permit a meaningful and up-to-date reinterpretation of speech rhythm. When the initial state of a dynamic deterministic system is known, its evolution is theoretically predictable. However, because of the complexity of systems, and in the absence of relevant predictive input, chaotic systems may rapidly evolve towards unpredictable states (Ruelle, 1989). The novelty in this view is related to the status given to the unknown. It changes the researcher's perspective. If we assume that speech temporal organisation is chaotic, then the unpredictable portion of calculated durations are no longer *random* values, but more likely *unexplained residuals*. Also, a system's chaotic nature sets up more realistic expectations about the limitations of our predictions, and it provides an intelligible account of when predictability will touch its limits (Kellert, 1993).

The comparison of syllabic durations in a reading task at two speech rates showed similar phase-plots, both converging toward a strange attractor but at different trajectories. The Hurst analysis suggested that syllables produced at fast speech rate show a greater difference from random motion (*i.e.*, more serial organisation) than those produced at a slow speech rate. Generally in our domain, speeding up speech is considered in terms of linear time compression. Zellner (1998) showed that in French, the problem is far more complex, and that the changing of speech rate involves non-linear transformations. The present chaotic analysis reinforces this view. Time modification in speech would thus have to be looked at with new eyes. Further investigation in this area will be examined in a forthcoming paper.

This article started with considerations on speech fluency. Fluent speech characterises an adapted temporal organisation, where sound and noise periods alternate with silent periods. Speech fluency favours the anticipation and adjustment of speech processing in such a way that it facilitates understanding (Hanes, 1986). Becoming fluent in a second-language is thus clearly an objective to be pursued by motivated students. In particular, they have to acquire the temporal organisation of the L2 system. Dysfluencies may occur, if the initial state of the temporal system, or any key succeeding state, is not properly synchronised. A program based on the acquisition of successively more complex temporal patterns in the segmental, syllabic, lexical and phrasal domains, embedded in the three levels of temporal speech constraints (bio-psychological, social and pragmatic constraints), would probably expose the student to the full range of complexity in this field, and would help improve the acquisition of temporal proficiency and fluency in the second language.

References

Allen, G. D. (1975). Speech rythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics, 3.* 75-86.

Astesano, C., Di Cristo, A. & Hirst, D.J. (1995). Discourse based empirical evidence for a multi-class accent system in French. *XIIIème Congrès International des Sciences Phonétiques, 4* (pp. 630-633). Stockholm.

Bailly, G. (1989). Integration of rhythmic and syntactic constraints in a model of generation of French prosody. *Speech Communication, 8*. 137-146.

Barbosa, P. A. (1994). *Caractérisation et génération automatique de la structuration rythmique du français.* Thèse de Doctorat. U.R.A. CNRS n°368 - INPG/ENSERG, Université Stendhal, Grenoble.

Beaugendre, F. (1994). Une étude perceptive de l'intonation du français. Thèse de Doctorat en Sciences de l'Université Paris XI. LIMSI n°94 - 25.

Berthoz, A. (1997). *Le sens du mouvement*. Paris: Odile Jacob.

Boë, L.J., Schwartz, J.L, & Vallée, N. (1992). la prédiction des structures sonores. *Programme Rhône-Alpes, Recherches en Sciences humaines, Rapport intermédiaire.*

Boissons-Bardies, B. (1996). Comment la parole vient aux enfants. Paris: O.Jacob.

Boomer, D.S., & Ditmann, A.T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech, 5,* 215.

Botte, M.C., Canévet, G., Demany, L., & Sorin, C. (1989). *Psychoacoustique et perception auditive.* Série audition. Paris: Inserm/ Sfa / CNET.

Bregman, A. S. (1994). L'analyse des scènes auditives: l'audition dans des environnements complexes. In S. McAdams, & E. Bigand (Eds), *Penser les sons. Psychologie cognitive de l'audition.* (pp. 11 - 39). Paris: P.U.F.

Caelen-Haumont, G. (1991). *Stratégies des locuteurs et consignes de lecture d'un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodiques,* Thèse d'Etat, Aix-en-Provence.

Chafe, W. (1980). Some reasons for hesitating. In W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 169-180). Mouton

Cook, M. Smith, J., & Lalljee, M. (1974). Filled pauses and syntactic complexity. *Language and Speech, 17,* 11-16.

Crutchfield, J., Farmer, J.D., Packard, N., Shaw, R. (1989)

Cutler, A. (1994). The perception of rhythm in language. *Cognition, 50*. 79-81.

Damasio, A. & Damasio, H. (1997). Le cerveau et le langage. *Pour la Science: Les langues du monde*. 8 - 15.

Dauer, R.M. (1983). Stress-timing and syllable timing reanalyzed. *Journal of Phonetics, 11*. 51-62.

Dauer, R. M. (1987). Phonetic and phonological components of language rythm. *XIème Congrès International des Sciences Phonétiques* (pp. 447-450). Talinn, Estonie. Proceedings.

Delais, E. (1994). Prédiction de la variabilité dans la distribution des accents et les découpages prosodiques en français. *20èmes Journées d'Etude sur la Parole* (pp. 379-384). Trégastel.

Delais, E, (1994), Rythme et structure prosodique en Francais, in Chantal Lyche (Ed.), *French Generative Phonology : Retrospectives and Perspectives*, , ESRI Publication, (pp. 131-150). UK.

Delais-Roussarie, E. (1995). *Pour une approche parallèle de la structure prosodique*: *étude de l'organisation prosodique et rythmique de la phrase française.* Thèse de Doctorat, Université de Toulouse-Le Mirail.

Di Cristo, A. & Hirst, D. (1994). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix, 15.* 13-24.

Duez, D., Nishinuma, Y. (1985). Le rythme en français. *Travaux de l'Institut de Phonétique d'Aix, 10*. 151-169.

Duez, D., & Nishinuma, Y. (1987). Vitesse d'élocution et durée des syllabes et de leurs constituants en français parlé. *Travaux de l'Institut de Phonétique d'Aix, 11*. 157-180.

Evans, J. R, Clynes, M. (1986). *Rhythm in psychological, linguistic, and musical processes*. Springfield, Charles C. Thomas.

Fant, G., Kruckenberg, A. (1996). Quantal theory of speech timing. *Speech, Music and Status Report. Fonetik 96*. 41-44.

Fónagy, I. (1992). Fonctions de la durée vocalique. In P. Martin (Ed.), *Mélanges Léon*. (pp. 141-164). Editions Mélodie-Toronto.

Fougeron, C. & Jun, S-A. (1995). Properties of french intonation at fast speech rate. *XIIIème Congrès International des Sciences Phonétiques, 3* (pp. 488-491). Stockholm.

Fraisse, P. (1974). *Psychologie du rythme*. Paris: Presses Universitaires de France.

Friedman, W. (1990). *About time. Inventing the fourth dimension.* Cambridge: MIT Press.

Gitta, P., Laan, M. (1992). Perceptual Differences between Spontaneous and Read Aloud Speech. *IFA Proceedings 16*. 65-79.

Goldman–Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* New York: Academic Press.

Goldman–Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*, 103-113.

Goldman, J. P., Content, A. & Frauenfelder, U. H. (1994). Comparaison des structures syllabiques en français et en anglais. *Actes des 20eme Journées d'Etude sur la Parole* (pp. 119-122). France: Lannion.

Gregory, R. L. (1993). *Le cerveau, un inconnu. Dictionnaire encyclopédique*. Paris: Robert Laffont.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica, 31.* 144-184.

Grosjean, F. , & Lane, H. (1981). Temporal variables in the perception and production of spoken and sign languages. In Eimas, P. D , & Miller, J. L (Eds). *Perspectives on the Study of Speech* (pp. 207-236). Hillsdale: Lawrence Erlbaum Associates.

Grover, C. & Terken, J. (1995). The role of stress and accent in the perception of rythm. *XIIIème Congrès International des Sciences Phonétiques, 4* (pp. 356-359). Stockholm.

Guaïtella, I. (1992). Rythme et parole: Comparaison critique du rythme de la lecture oralisée et de la personne spontanée,. *Travaux de l'Institut de Phonétique d'Aix,* 14, 221-225.

Guaïtella, I. (1996). Parole spontanée et lecture oralisée: activités cognitives différentes, organisations rythmiques différentes. *Travaux de l'Institut de Phonétique d'Aix, 17.* 13-30.

Hanes, M. (1986) in J. R, Evans, & M. Clynes, (Eds), *Rhythm in psychological, linguistic, and musical processes* (pp. 99 - 130). Springfield, Charles C. Thomas.

Hausdorff, JM., Peng, CK., Ladin Z, Wei, JY., Goldberger, AL. (1995). Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *Journal of Applied Physiology, 78(1)*:349-58

Hirst, D. & Di Cristo, A. (Eds.). (1998). *Intonation Systems: a Survey of Twenty Languages.* Cambridge University Press

Jun, S-A. & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. *XIIIème Congrès International des Sciences Phonétiques, 2* (pp. 722-725). Stockholm.

Keller, E. (1990). Speech Motor Timing. In W.J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* [pp. 343-364]. Amsterdam: Kluwer.

Keller, E., Zellner, B., Werner, S., and Blanchoud, N. (1993). The prediction of prosodic timing: Rules for final syllable lenthening in French. *Proceedings ESCA Workshop on Prosody, September 27-29. Lund, Sweden.* 212-215.

Keller, E., & Zellner, B. (1995). A statistical timing model for French. *XIIIème Congrès International des Sciences Phonétiques, 3* (pp. 302-305). Stockholm.

Keller, E., & Zellner, B. (1996). A timing model for fast French. *York Papers in Linguistics*, 17, University of York. 53-75.

Kellert, S. (1993). *In the Wake of Chaos.* The University of Chicago Press.

Kelso, J. S., Saltzman, E. & Tuller, B. (1986). The dynamical perspective in speech production: data and theory. *Journal of Phonetics, 14.* 29 - 59.

Ladd, R., & Campbell, N. (1991). Theories of prosodic structure: evidence from syllable duration. *XIIIème Congrès International des Sciences Phonétiques*, *2.*(pp. 290-293). Aix en Provence.

Lehiste, I.(1972). The timing of utterances and linguistic boundaries.*Journal.of the Acoustical Society of America*,*69.* 2018–2024.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5.* 253-263.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. MIT Press. Cambridge.

Liberman, A.M., Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H.W. Leibowitz & H.L. Teuber (Eds.), *Handbook of sensory physiology*, vol 8 (pp. 431-461). New-York: Sprenger-Verlag.

Liberman, A.M., Mattingly, I.G. (1985). The motor theory of speech perception revisited. *Cognition* 21, 1-36.

Lindblom, B. (1996). Approche intégrée de la production et de la perception. *Fondements et perspectives en traitement automatique de la parole. (pp. 9 - 17).* AUPELF-UREF.

Martin, P. (1987). Structure rythmique de la phrase française. Statut théorique et données expérimentales. *Actes des 16e Journées d'Etude sur la Parole* (pp. 255-257). *Hammamet.*

McAdams, S. & Bigand, E. (1994). *Penser les sons. Psychologie cognitive de l'audition.* Paris: P.U.F.

Mertens, P. (1990). Intonation. In Blanche-Benveniste, C., & al (Eds), *Le français parlé. Etudes grammaticales* (pp. 159-176). Paris: Editions du CNRS.

Miller, N., Maruyana, G., Beaber, R.J. & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology, 4.* 615-624.

Newmeyer, F. J. (1988). *Linguistics: The Cambridge survey.* Vol IV. Cambridge University Press.

Nishinuma, Y, & Duez, D. (1988). Etude perceptive de l'organisation temporelle de l'énoncé en français. *Travaux de l'Institut de Phonétique d'Aix, 11.* 181-201.

Nishinuma, Y, & Santi, S. (1992). Comment percevons-nous la dernière syllabe d'une phrase? Relations entre intensité et durée dans la discrimination de segments vocaliques. *Travaux de l'Institut de Phonétique d'Aix, 14.* 149-16.

Pasdeloup, V. (1990). Organisation de l'énoncé en phases temporelles: Analyse d'un corpus de phrases réitérées. *Actes des 18émes Journées d'Etudes sur la Parole.*(pp. 254 - 258). Montréal.

Pasdeloup, V. (1992.a). A prosodic model for French text-to-speech synthesis: A psycholinguistic approach. In G. Bailly, & al (Eds.), *Talking Machines. Theories, Models, and Designs* (pp. 335-348). Elsevier Science Publishers.

Pasdeloup, V. (1992.b). Durée intersyllabique dans le groupe accentuel en français. *Actes des 19émes Journées d'Etudes sur la Parole*. (pp. 531-536). Bruxelles.
Patel, A.D., Peretz, I., Tramo, M., Labreque, R. (1998). Processing prosodic and musical patterns: a neurophysiological investigation. *Brain and Language, 61*, 123 - 144.
Pfauwadel, M.-C. (1986). *Etre bègue.* Paris: Retz.
Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
Robert, L. (1989). *Les horloges biologiques*. Paris: Flammarion.
Rosenfield, D. (1989). A neuroscience model of stuttering. *Journal of Fluency and Disorders*. *14.*
Ruelle--
Scott, S. K. (1993). *P-centres in speech: an acoustic analysis.* PhD Thesis, University College London.
Scherer, K. (1984). Les émotions: fonctions et composantes. *Cahiers de Psychologie Cognitive, 4*. 9-39.
Scherer, K. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin, 99*. 143-165.
Scherer, K. & Zei, B. (1989). La voix comme indice affectif. *Revue médicale de la Suisse Romande,* 109. 61-66.
Sendlmeier WF (1995). Feature, phoneme, syllable or word: how is speech mentally represented? *Phonetica*. 52(3):131-43, 1995.
Starkweather, W.C. (1987). *Fluency and Stuttering,* Prentice Hall.
Stevens K. (1989). On the Quantal Nature of Speech. Journal of Phonetics, 17. 3-45
Street, R. & Brady, R. (1982). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs, 49*. 290-308.
Vaissière, J. (1991). Rhythm, accentuation and final lengthening, in J. Sundberg, L. Nord & R. Carlson (Eds.), *French in Music, Language, Speech and Brain* (pp.108-120). Wenner-Gren International Symposium Series Macmillan Press, Vol. 59.
Vincent, M., Di Cristo, A. & Hirst, D. (1995). Prosodic features of finality for intonation units in French discourse. *XIIIème Congrès International des Sciences Phonétiques, 2* (pp. 718-721). Stockholm.
Wang, M. Q. & Hirchberg, J. (1992). Automatic Classification of Intonational Phrase Boundaries. *Computer Speech and Language,6*. 175-196.
Warren, R. M. (1994). La perception des séquences acoustiques: intégration globale ou résolution temporelle? In S. McAdams, & E. Bigand (Eds), *Penser les sons. Psychologie cognitive de l'audition.* (pp. 41 - 72). Paris: P.U.F.
Wenk, B. J., & Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10.* 177-193.

Wightman, C. W. (1992.a). *Automatic detection of prosodic constituents for parsing.* Doctoral dissertation. Boston University Graduate School.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992.b). Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *J. Acoust. Soc. Am.* 91, 1707-1717.

Wiolland, F. (1984). Organisation temporelle des structures rythmiques du français parlé. Etude d'un cas. *Actes des Rencontres régionales de Linguistique, BLLL* (pp. 293 - 322).

Zellner, B. (1992). Le bé bégayage et euh... l'hésitation en français spontané. *Actes des 19ème Journées d'Etudes sur la Parole* (pp. 481-487). Bruxelles.

Zellner, B. (1994). Pauses and the temporal structure of speech, in E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition.* (pp. 41-62). Chichester: John Wiley.

Zellner, B. (1996). Structures temporelles et structures prosodiques en français lu. *Revue Française de Linguistique Appliquée: La communication parlée.* 1.Paris. 7-23.

Zellner, B. (1997). Improving Speech Fluency in French through Psycholinguistic Principles. *14th CALICO Annual Symposium,* (CD) ISBN 1-890127-01-9. New-York.

Zellner, B. (1997). Fluidité en synthèse de la parole. In E. Keller, & B. Zellner (Eds.), *Les défis actuels en synthèse de la parole*e, *Études des Lettres, 3*. (pp. 47-78). Université de Lausanne.

Zellner, B. (1998). *Caractérisation et prédiction du débit de parole en français. Une étude de cas.* Thèse de Doctorat. Faculté des Lettres, Université de Lausanne.

Zwicker, E., & Feldkeller, R. (1981). *Psychoacoustique: l'oreille récepteur d'information.* Collection technique et scientifique des télécommuications. Paris: Masson.