Silent reading and prosodic structure constraints

Philippe Martin

UMR 7110, LLF, UFRL, Université Paris Diderot, ODG, Place Paul Ricœur, 75013 Paris, France
philippe.martin@linguist.univ-paris-diderot.fr

Abstract

Silent reading of written texts involves normally a process of subvocalization, i.e. the presence of a voice reading the text in the head of the reader speaking to her/himself. This process includes not only the sequences of syllables corresponding to the written material, but also sentence intonation. Since subvocalization cannot be eliminated other than by changing the status of each word into a pictographic function (as it may be the case for a STOP road panel sign for example), it is argued here that sentence intonation is essential to language comprehension, and more specifically to the conversion of sequences of syllables into higher order linguistic units (corresponding to accent phrases AP in the Autosegmental-Metrical model).

Consequently, reading and in particular silent reading is constrained by the same rules than the prosodic structure in general, and specifically to the minimal duration of accent phrases. This minimal value, occurring when AP's contain only one syllable, is about 250 ms, a value which corresponds to the minimal period value of Delta brain waves [4], [11]. Therefore, this AP minimal duration limits also the maximal number of AP that could be processed in silent reading, i.e. about 240 per minute, which corresponds to the maximal number of words per minutes experts in fast reading can process while keeping a reasonable level of comprehension, i.e. about 800 wpm.

Index Terms: silent reading, prosodic structure, subvocalization, Delta waves, Theta waves.

1. Introduction

When we read, either silently or aloud, we generate speech sounds according to the reduced information given in the written text. In this process, we also generate a prosodic structure, which hierarchically organizes accent phrases AP (minimal units of intonation containing a single lexical or group stress), into prosodic groups, called in the Autosegmental-Metrical model ip (intermediate intonative phrases) and at a higher level IP (Intonation Phrases), whose sequences in turn constitute the whole utterance intonation.

It is noticeable that this prosodic structure (re)generation is essential to help the reader to understand the text. Therefore, the whole reading process is constrained by the rules governing the elaboration of the sentence prosodic structure when speaking either silently or aloud, and in particular the minimal and maximal duration of accent phrases [11].

In silent reading, it is difficult to proceed without subvocalization, i.e. without hearing a voice in one's head that corresponds to a voice reading the text aloud. For this reason, silent reading may be subject to similar constraints than reading aloud (let aside articulatory constraints). These constraints may interact or even supersede constraints established for eye movement while reading. In particular, they may lead to a new explanation pertaining to the maximum number of words that can be processed in fast reading.

2. Eye movement

When reading, the eye proceeds in saccades (short rapid movements) to scan the text, jumping in steps varying from 1 to 20 characters with an average of 7 to 9 characters (forward and backward). In the process, the most frequent fixations are given by verbal forms and punctuation marks (dots, commas, semicolons, question marks, etc.). The eye jumps then constantly to spot these markers, which will constitute the bases for the prosodic structure to build [10].

Most of the laboratory speech research on sentence intonation actually investigate this process thoroughly on read speech, before considering spontaneous, non-prepared speech prosodic features. For example, if a dot normally ending written text sentences is associated with a falling conclusive prosodic marker, the correspondence of the other punctuation marks and the verbal forms must be dynamically associated with a proper prosodic contour (a Tone Boundary in the Autosegmental-Metrical model).

The saccades allows to eye to focus on fixation point in 20 ms to 40 ms, whereas the fixation point last between 100 ms and 500 ms [17]. The fixation state of the eye allows the fovea, the central part of the retina, to scan the selected written information with high resolution, whereas peripheral information is viewed with fewer details (Fig. 1).

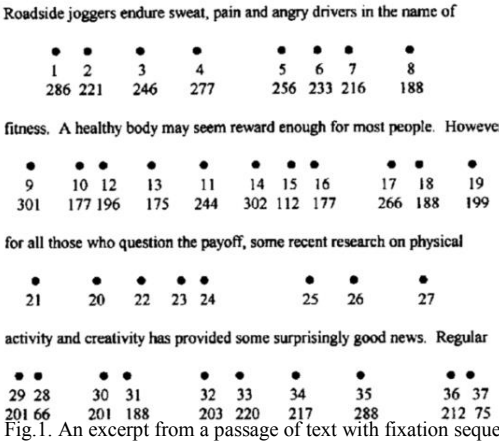

Fig.1. An excerpt from a passage of text with fixation sequence and fixation durations. The dots below the words indicate the fixation location, the first number below a dot its rank in the sequence, and the second number below a dot its duration of fixation (from [17]).

Due to the complex muscular mechanisms for speech generation, oral (i.e. aloud) reading is slower than silent reading. However, the puzzling aspect of silent reading lies in its limitations. Despite a number of questionable commercial claims stating that fast readers could read up to 3000 words per minute (about 50 words/sec…), the fast reading process is limited by subvocalization, the effect to hear a voice in one's head while reading silently (which was curiously attributed by some to the way we learn to read at school [13]).

10.21437/SpeechProsody.2014-144

3. Subvocalization

Subvocalization does not pertain to the mechanical control of articulators muscle control, but to the perception of the speech signal, which is recovered by reading. The invention of writing has precisely this function, allowing not only reading aloud but also silently, i.e. "to talk to oneself in silence". Indeed, writing is a shorthand notation system of speech sound and not of articulatory movements, contrary to what claim supporters of the motor theory of speech perception [8]. There is apparently no writing system (even API) referring directly to articulatory configurations.

Other systems such as pictograms bypass the generation of speech sounds by associating directly significant and signifier to access their signification without going through language units be syllables, words, prosodic words, syntagms, etc. A road STOP sign may indeed be read aloud or silently, but is more frequently directly associated with its meaning, i.e. to give way on the road.

Likewise, dates written with numbers, e.g. 1789, may be read as "seventeen hundred eighty nine", but the constant use of symbols not corresponding directly to syllables and words leads more frequently to a direct access to its signifier. The passage to the status of pictogram depends of course of the familiarity of the reader with the object and its frequency of occurrence.

Writing systems using ideograms, for example Mandarin, also involve subvocalization in silent reading. Learning Mandarin without being concerned by ideograms pronunciation would be difficult, as many words are plurisyllabic, implying for such reader to deal with combination of pictograms [9]. However, one could associate other sounds to ideograms, such as English words for example, but the mediation of some speech sound seems difficult to avoid, although not impossible in principle.

Commercial US based fast reading "schools" claim that they can remove subvocalization, or at least minimize it. The subliminal idea is to transform every word into a pictogram, so when read it will not be pronounced silently. Other techniques recommend to use a pencil to determine eye fixation targets and accelerate the number of saccades. An application even proposes to display only lexical words sequentially on a computer screen with a user adjustable speed (this approach incidentally corresponds to the definition of accent phrases in the autosegmental-metrical model, i.e. one lexical word for each accent phrase). Comprehension should then be achieved without any prosodic structure and no syntactic structure linking the read words together by simple concatenation and no hierarchical structure.

Faster readers claim speed from 400 wpm (words per minutes) to 800 wpm. With an average number of about 3 (written) words per accent phrase, 800 wpm convert into about 266 AP's per minute, or 266/60 = 4.4 AP's per second. So the minimal average duration between silently read AP's would be about 225 ms. 800 wpm for the best observed performance for speed readers [2].

4. Accent phrases

Recent studies on spontaneous speech show that we speak and read by accent phrases and not word by word [1]. Accent phrases (AP's, aka prosodic word, stress groups, temporal groups, etc.) are minimal units of prosody contain only one (lexical or group) stress. It is also claimed that accent phrases necessarily contain either a verb, a noun, an adjective or an adverb together with grammatical words, but the analysis of non-prepared speech (i.e. spontaneous) data invalidated this hypothesis [12].

Accent phrase duration measured on various styles of spontaneous speech show that the shortest values, corresponding to a single syllable accent phrase, is about 250 ms, even if the single syllable is reduced to a single vowel, which would otherwise take some 100 ms to 150 ms when unstressed [12]. This minimal duration seems to be a limit under which the syllable ceased to be perceived as prominent (i.e. stressed).

The longest duration of accent phrase is about 1200 ms, which corresponds to an average of 7 syllables. This implies that sequences of more than 7 syllables or so must contain more than one stressed syllable, as in the English word *paraskevidekatriafobia* (the fear of Friday 13) realized with two or three stressed syllables: *paraske'**vi**dekatriafo'**bi**a* or *pa'raskevide'katriafo'**bi**a* for example. The question is: why do we have these lower and upper duration limits for accent phrases?

5. Theta and Delta brain waves

In the years 1930-40, researchers observed that the human brain consisted of a very large number of neurons (in the order of 100 billions) interconnected in groups in specific regions of the brain mass. These interconnections allow a transfer of chemically stored information in each neuron. These transfers induce variations of a small electric potential (in the µV range), that can be observed through captors positioned on subjects skull (electroencephalography, or EEG). These electrical variations are called evoked potential as they result from a sensory stimulation, auditory, visual or other.

Electrical activity produced by transfers of group of neurons to other groups of neurons is not done haphazardly. First, they operate in specific frequency ranges linked to specific cognitive activities, and secondly they can be synchronized in phase in each frequency range. Greek letters designate specific frequency ranges: Alpha, Delta, Delta, Gamma… The range of interest here are Delta, varying from 1 to 4 Hz, and Theta, varying from 4 to 10 Hz.

Evoked potential is usually observed with a relatively large number of captors (from 32 to more than 256 today) placed around the subject skull according to location standards. EEG signals are stored in real time and analyzed into the frequency domain with either a (Fast) Fourier or Wavelet transform. The resulting representation is very similar to spectra obtained in frequency speech analysis, the frequency ranges being of course much lower.

The constraints governing temporal groups observed on prosodic structures of both read and spontaneous speech can find a justification – and an explanation – in recent neurophysiological research work on speech ([4], [5], [11]). These studies, based essentially on EEG, investigate the possible correlations that may exist between brain activity and the perception and linguistic treatment of the information by listeners.

Researchers in neurolinguistic for instance, demonstrated with this technique of investigation the precedence of prosodic over syntax treatment [19]. Experiments described in [14] and [4] showed that the speech flow was segmented thanks to prosodic tags and with direct identification of already

memorized units. Two complementary processes would explain the conversion of syllabic flow into higher order units, which would result preferential lateralization to the right hemisphere for the prosodic information, and the left hemisphere for language information already stored.

Following proposals put forward in [3], [4] and [5], these observations lead to the following hypothesis. We know that the waves of the cortex Theta and Delta (among others) govern the flow of information transfer between neuronal groups. Delta wave frequencies ranging from 1 to 4 Hz, while those of Theta waves range from 4 to 10 Hz. These values suggest that Delta waves are responsible for the timing of the transfer of syllabic sequences, the syllables storage in short-term memory being synchronized by Theta waves.

Figures 2 and 3 below illustrate the difference pertaining to EEG recordings resulting from a stimulus of unstructured pure tones (Fig. 1) and a structured sequence organized in 4 chunks of 3 pure tones, the last tone of each chunk with a longer duration (Fig. 2).

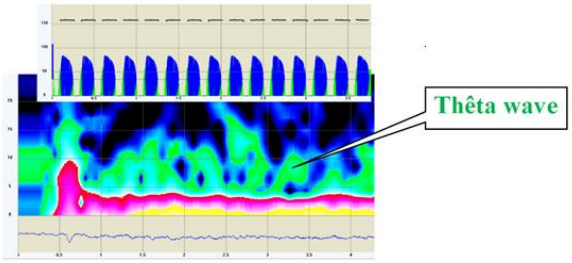

Fig. 2. Example of EEG spectral analysis (channel 28 or Pz) of evoked potential for a stimulus of a sequence of pure tones (top of the figure). Spectral analysis (bottom) shows Theta waves (in the range 4 Hz-10Hz) with no temporal structure [11].

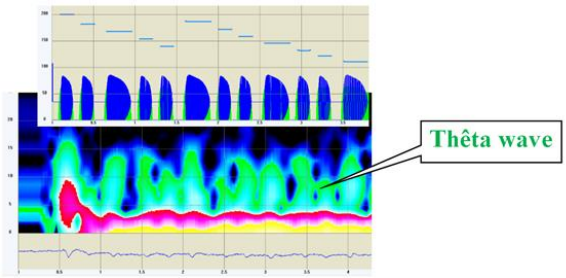

Fig. 3. Example of EEG spectral analysis (channel 28 or Pz) of evoked potential for a stimulus of a sequence of pure tones structured in groups of 3, the third tone being longer (top of the figure). Spectral analysis (bottom) shows Theta waves (in the range 4 Hz-10Hz) organized in a temporal structure corresponding to the stimulus structure [11].

Fig. 2 and 3 above suggest that temporal groups can only be perceived by the listener if their conversion is triggered by Delta waves (which may also synchronize Theta waves). This process is therefore constrained by the Delta wave properties, and in particular by its frequency properties. This hypothesis would account for 1) the extent of variation of the durations of stress groups, ranging from 250 ms to about 1200 ms (variation range of wave periods Delta) and 2) variation periods of Theta waves, from 100 ms to 250 ms.

6. Conclusion

Delta brain waves, whose periods vary from 250 ms to 1200 ms (about 1 Hz to 4 Hz), synchronize the conversion of sequences of syllables stored in short-term memory into higher linguistic units. This hypothesis is validated by the minimal and maximal duration of accent phrases (stress groups) which correspond to the Delta period variations. Furthermore, the average duration of syllables decreases linearly with their number in an accent phrase, from about 250 ms to 100 ms in accent groups of 1 to 7 syllables [11].

As no actual acoustical speech production is involved, silent reading is much faster than reading aloud, where multiple muscular command must be executed. Still, although eye saccade and eye fixation can operate much faster, for example in the reading of pictograms, subvocalization essential in silent reading limits the reading speed. Indeed, since subvocalization implies the generation of sentence prosodic structures as well as sequences of syllables, a prosodic constraint limiting the minimum duration of accent phrase to about 250 ms limits also the speed of the silent reading process, which has to go necessarily through this prosodic structure regeneration process. These values correspond tightly to the fastest reading performances cited in the literature [2], i.e. about 800 wpm or 4 AP per second.

7. References

[1] Blanche-Benveniste, Claire (2003) La naissance des syntagmes dans les hésitations et répétitions du parler, in Araoui J.L. ed. *Le sens et la mesure*. *Hommages à Benoît de Cornulier*, Honoré Champion, Paris, 2003, 40-55.

[2] Dunning, Brian (2010) Speed Reading, *Skeptoid Podcast*. Skeptoid Media, Inc., 26 Oct 2010. Web. 12 Dec 2013. http://skeptoid.com/episodes/4229

[3] Friederici, Angela & Wartenburger, Isabell, 2010, Language and brain, *Cognitive Science*, (10) 150-159.

[4] Ghitza1, Oded, Giraud, Anne-Lise and Poeppel, David (2013) Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence, *Frontiers in Human Neuroscience*, www.frontiersin.org, January2013, Volume 6, Article340.

[5] Gilbert, Annie & Boucher, Victor (2007) What do listeners attend to in hearing prosodic structures? Investigating the human speech-parser using short-term recall, *Proc*. *Interspeech 2007*: 430-433.

[6] Giraud, Anne-Lise and David Poeppel (2012) *Cortical oscillations and speech processing: emerging computational principles*. Nature neuroscience E-pub, doi: 10.1038/nn.3063.

[7] Just, Marcel Adam and Patricia A. Carpenter (1987) *The Psychology of Reading and Language Comprehension*. Boston: Allyn and Bacon, 1987.

[8] Liberman, Alvin M. and Ignatius G. Mattingly (1985) The motor theory of speech perception revised, *Cognition* 21 (1): 1–36

[9] Marshall Unger, James (2003) *Ideogram: Chinese Characters and the Myth of Disembodied Meaning*, University of Hawai'i Press, 216 p.

[10] Martin, Philippe (2011) *Ponctuation et structure prosodique*, Langue Française, 2011, nº 172, 99-114.

[11] Martin, Philippe (2013) Contraintes phonologiques de l'intonation de la phrase réinterprétées à la lumière des recherches récentes en neurophysiologie, *La Linguistique*, 2013/1.

[12] Martin, Philippe (2014) Spontaneous speech corpus data validates prosodic constraints, submitted to *Speech Prosody 2014 Conference*, Dublin 2014.

[13] Nowak, Paul (2012) *Speed reading tips: 5 ways to minimize subvocalization*, http://www.irisreading.com/speed-reading/speed-reading-tips-5-ways-to-minimize-subvocalization/

[14] Obrig, Hellmuth, Rossi, Simone, Telkemeyer, Silke & Watenburger, Isabell, 2010, From acoustic segmentation to

language processing: evidence from optical imaging, *Front. Neuroenerg*, 2:13.
[15] Rayner, Keith and Alexander Pollatsek (1989) The Psychology of Reading, Lawrence Erlbaum Associates, Hillsdale, NJ, 544 p.
[16] Rayner, Keith, Barbara Foorman, Charles A. Perfetti, David Pesetsky, and Mark S. Seidenberg (2001) How Psychological Science Informs the Teaching of Reading. *Psychological Science in the Public Interes*t **2** (2): 31-74.
[17] Reichle, Erik D., Pollatsek, Alexander, Fisher, Donald L. and Keith Rayner (1998) Toward a Model of Eye Movement Control in Reading, *Psychological Review* 1998, Vol. 105, No. 1, 125-157
[18] Sereno, Sara and Keith Rayner (2003) Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Science* **7** (11): 489 - 493.
[19] Steinhauer, Karsten, Alter, Kai & Friedrici & Angela D., 1999, Brain potentials indicate immediate use of prosodic cues in natural speech processing, *Nature Neuroscience*, 2(2) 191-196.